(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(54) Title: MALARIA IMMUNOGEN AND VACCINE

(57) Abstract: A chimeric, carboxy-terminal truncated hepatitis B virus nucleocapsid protein (HBc) is disclosed that contains an immunogen for inducing the production of antibodies to malarial proteins. An immunogenic malarial epitope is expressed between residues 78 and 79 of the HBc immunogenic loop sequence. The chimer preferably contains a malaria-specific T cell epitope and is preferably engineered for both enhanced stability of self-assembled particles and enhanced yield of those chimeric particles. Methods of making and using the chimers are also disclosed.

MALARIA IMMUNOGEN AND VACCINE

Description

TECHNICAL FIELD

The present invention relates to the intersection of the fields of immunology and protein engineering, and particularly to an immunogen and vaccine useful in prevention of malaria infection by *P. falciparum* or *P. vivax*.

BACKGROUND OF THE INVENTION

Malaria is by far the world's most important tropical parasitic disease, killing more people than any other communicable disease, with the exception of tuberculosis. The causative agents in humans are four species of *Plasmodium* protozoa: *P. falciparum, P. vivax, P. ovale* and *P. malariae*. Although *P. falciparum* accounts for the majority of infections and is responsible for the vast majority of deaths attributable to malaria, *P. vivax* causes a recurring chronic debilitating disease for which a vaccine is necessary.

Malaria infection begins when a female *Anopheles* mosquito infected with one of the four *Plasmodium* species infectious for humans bites a person. The mosquito's saliva carries the malarial sporozoites into the blood. Approximately 30 minutes later these sporozoites enter the liver. Once in the liver, the sporozoites divide over the

course of about 5 days, forming a schizont. A schizont
may contain up to 30,000 merozoites, which spill into the
bloodstream when the schizont ruptures.   Within seconds,
merozoites infect red blood cells (RBCs) and again
replicate asexually, with each schizont producing up to 36
merozoites.

Each time a RBC bursts and liberates progeny,
other blood cells are infected. The cycle can continue
until the person dies of anemia and/or other
complications. A few of the merozoites in RBCs
differentiate into gametocytes, a sexual form, which, if
ingested by a mosquito, are liberated from the RBCs in the
mosquito stomach and subsequently mate. The progeny,
sporozoites, accumulate in the saliva and the process
starts again when the mosquito feeds [See, Hoffman et al.,
(1996) "Attacking the Infected Hepatocyte", in *Malaria
Vaccine Development* (ed. S.L. Hoffman), p.35. ASM Press,
Washington, D.C., for review).

P. vivax malaria is most prevalent in Latin
America (where it is as or more prevalent than *P.
falciparum*) and Asia. Although rarely fatal, *P. vivax*
malaria has a dormant liver phase that is associated with
relapses that show variability in duration, depending on
the strain.

Presently, there is not an effective vaccine
against any form of malaria. For many years, chloroquine
was a cheap and effective therapeutic for treating
malaria. But in recent years, chloroquine resistance has
increased dramatically for *P. falciparum*. In the past 10
years, *P. vivax* has also developed chloroquine resistance
with cases being reported in south-east Asia, the south-
west Pacific; Burma [Marlar, T., et al., *Trans. R. Soc.
Trop. Med. Hyg.*, 1995. 89(3): p. 307-308] perhaps India
[Garg et al., Trans. R. Soc Trop. Med. Hyg., 1995. 89(6):

p. 656-657],Indonesia [Baird et al., *Am. J. Trop. Med. Hyg.*, 1991, 44(5): p. 547-552; Baird et al., *Trans R Soc Trop Med Hyg*, 1996, 90(4): p. 409-411; Baird et al., J. Infect. Dis., (1995) 171(6): p. 1678-1682; Murphy et al., *Lancet*, (1993) 341(8837): p. 96-100; and Schwartz et al., [letter]. N. Engl. J. Med., (1991) 324(13): p. 927] and Papua New Guinea [Rieckmann et al., *Lancet*, (1989) 2(8673): p. 1183-1184].

Primaquine is the only antimalarial drug that is effective against hyponozoites, which are associated with the dormant phase in the liver responsible for relapses. Different *P. vivax* strains show differential patterns of relapse; for example, a Korean *P. vivax* strain has been shown to be 100 percent radically cured by a given primaquine regime (WHO, 1967), whereas the same regime is only 70 percent effective with the Chesson strain [Coatney et al., *J. Natl. Malaria Soc.*, 1962. 9: p. 285-292]. To complicate treatment with primaquine further, reports in the 1970s highlighted primaquine-resistant P.vivax in south-east Asia [Charoenlarp et al., *Southeast Asian J. Trop. Med. Public Health*, (1973) 4(1): p. 135-137 and Krotoski, *[letter]*. N. Engl. J. Med., (1980) 303(10): p. 587]; observations have steadily increased in other locations in recent years [Schuurkamp et al., *Trans. R. Soc. Trop. Med. Hyg.*, (1992) 86(2): p. 121-122] suggesting that widespread resistance to primaquine is emerging.

Clearly, the most effective approach to combating malaria is an effective vaccine. As with smallpox, and potentially polio in the near future, a coordinated worldwide vaccination program can result in eradication of communicable human diseases. This may also be achievable for malaria if an effective vaccine can be developed.

There are three recognized anti-parasitic approaches to malaria vaccine development. These are proposed to function by interrupting the parasite's lifecycle at three different stages.

The first and most attractive approach is the pre-erythrocytic vaccine, which aims to block sporozoite entry into the hepatocyte and/or release of merozoites into the blood stream. Immediately following infection, sporozoites migrate to the liver and begin the exoerythrocytic stage of their lifecycle. Successful blocking of hepatocyte entry, or the destruction of infected hepatocytes prior to liberation of merozoites, would prevent the disease, the passage of the parasite on to feeding mosquitoes, and merozoite release and subsequent invasion of red blood cells.

A second approach is to develop an 'antidisease' vaccine. The target is the red blood cell stage of the infection, during which the parasite grows at an exponential rate. Also known as 'asexual blood stage' vaccines, the merozoite surface protein 1 (MSP-1) and apical membrane antigen 1 (AMA-1) protein have emerged as the two most promising vaccine candidates for intervening at this stage of the disease (See, Good, et al., *Anu. Rev. Immunol.*, (1998) 16: p. 57-87,, for review). This stage is thought to represent a conceptually more difficult target compared with the pre-erythrocytic stage, which is associated with 10-20 sporozoites per mosquito bite, due to the tremendous increase in parasite load once the blood stage is reached.

A third approach, known as the 'transmission-blocking' vaccine, would not stop infection or symptoms in the individual. However, it would prevent infection from spreading to others by blocking the lifecycle in the mosquito by inducing antibodies that the mosquito would

-4-

ingest from the host with its blood meal. This vaccine
approach is more attractive as a long-term global solution
to eradication of malaria and less attractive to the
immediate needs of travelers and the military forces.

In the 1960s, researchers at New York University
(NYU) achieved full protection from malaria infection by
injecting animals with small numbers of sporozoites from
mosquitoes that had previously been irradiated. Later,
researchers at the University of Maryland, NYU and Walter
Reed Army Institute showed that percent of a group of
human volunteers immunized with irradiated sporozoites
later resisted exposure to virulent sporozoites [Clyde et
al., Am. J. Med. Sci., (1973) 266(6): p. 398-403 and
Rieckmann et al., Trans. R. Soc. Trop. Med. Hyg., (1974)
68(3): p. 258-259]. This work confirmed that protective
immunity to the sporozoite stage (i.e. the pre-
erythrocytic stage) of the malaria parasite could be
induced. However, an inability to culture sporozoites in
vitro thwarted the possibility of using them as a vaccine.

The strategic development of a synthetic malaria
vaccine required the identification of immunodominant,
neutralizing malaria epitopes. In 1985, a group at NYU
led by Drs. Ruth and Victor Nussenzweig, identified the
dominant B cell epitope from the circumsporozoite protein
(CS), a major component of the sporozoite surface membrane
at the time the parasite enters the bloodstream [Zavala et
al., Science, (1985) 228(4706): p. 1436-40]. Antibodies
to the repeated epitope were shown to be sporozoite
neutralizing by protecting against rodent and human
malaria [Nussenzweig et al., Ciba Found. Symp., (1986)
119: p. 150-163]. Antibodies to the CS protein also
correlated positively with protection in immunized mice
and in naturally infected individuals.

These studies strongly suggest that anti-CS repeat antibodies alone are able to confer protection against malaria infection, provided sufficient antibody titers can be raised. The identification of this epitope therefore enabled, for the first time, the strategic development of synthetic CS-based malaria vaccines. Several malaria vaccine candidates employing different carriers were developed based upon the identification of this epitope. The main focus of malaria vaccine development has been on *P. falciparum*, and it is widely assumed that information gained from studying *P. falciparum* extend to other *Plasmodium* species, including *P. vivax*. A brief overview of four pre-erythrocytic *P. falciparum* malaria vaccine candidates is given below.

The (NANP)$_3$ synthetic peptide conjugated to the protein carrier tetanus toxoid (TT) was the first synthetic malaria vaccine to undergo phase I and phase II clinical trials in the late 1980s [Etlinger et al., *Immunology*, (1988) **64**(3): p. 551-558; Etlinger et al., *J. Immunol.*, (1988) **140**(2): p. 626-633 and Herrington et al., *Nature*, (1987) **328**(6127): p. 257-259]. TT is widely known to provide powerful T cell help for coupled immunogens. Of the thirty-five vaccinees, the three having the highest titers of anti-sporozoite antibodies were selected for challenge studies. One of the vaccine recipients remained free of parasitaemia at 29 days, whereas the other two did not exhibit asexual stage parasites until 11 days, compared with a mean of 8.5 days for the un-vaccinated control group. Therefore, protection again correlated positively with anti-NANP titers.

The limited effectiveness of this vaccine was attributed to suboptimal levels of anti-NANP antibodies. Attempts to increase dosage were hindered by toxicity of the TT carrier. Further, the lack of parasite-derived

determinants capable of priming malaria-specific T cells also likely contributed to the low levels of protection.

Short synthetic peptides often have an *in vivo* half-life that is too short for them to be effective as prophylactic or therapeutic drugs. Standard approaches for increasing the immunogenicity of peptides is to either couple them to larger carrier proteins, or to assemble them into multimeric structures. In this case, 32 copies of the CS repeat sequence $((NANP)_{15}(NVDP))_2$ were linked and recombinantly fused to a random 32 amino acid fusion protein 20. [Ballou et al., *Lancet*, (1987) 1(8545): p. 1277-1281.] This vaccine candidate was called FSV-1.

Following immunization, twelve of the fifteen volunteers developed antibodies that reacted with sporozoites. No patients exhibited adverse reactions to the protein, indicating that the NANP (SEQ ID NO: 184) repeat itself is non-toxic. Of the fifteen patients immunized with 3 doses, six were selected to receive a fourth dose and were then challenged with the malaria parasite. Parasitaemia did not develop in the volunteer with the highest titer of CS antibodies, and parasitaemia was delayed in two of the other five vaccinees.

As with the NANP-TT vaccine discussed above, protection correlated positively with anti-NANP titers. This vaccine was deemed partially successful in that it reconfirmed that humans can be protected by CS protein subunit vaccines. However, the level of protection was not sufficient to warrant larger trials of this particular candidate.

A major shortfall of this vaccine was that it did not provide an efficient source of T cell help. The only individuals who would have received T cell help from this vaccine would be those in whom the CS repeat served as both a B and T helper (Th) cell epitope. However, this

sequence is known to be a Th epitope for only a limited number of individuals; i.e. it is highly genetically restricted.

Nardin and coworkers at NYU have been able elicit relatively high titers of anti-CS antibody in a diverse range of genetic backgrounds by combining the NANP repeat epitope with the T cell site identified by Berzofsky and Good [Good et al., *Science*, (1987) 235(4792): p. 1059-62] in a MAP format [Calvo-Calle et al., *J. Immunol.*, (1993) 150(4): p. 1403-1412]. Using their proprietary 'universal' form of the CS T cell epitope, Nardin and co-workers have been able to elicit anti-CS antibodies in all genetic backgrounds tested, suggesting that genetic restriction is alleviated by inclusion of this epitope.

Although MAPs have proven to be excellent research tools, providing valuable insight into immune recognition of the CS protein, there are several intrinsic problems associated with using them in a commercial vaccine. Their commercial utility has yet to be established relative to manufacturing and cost issues. Nevertheless, ongoing human clinical testing of these vaccine candidates will provide very useful information pertaining to the actual anti-NANP titers necessary for protective immunity.

One of the most promising malaria vaccines of recent times utilizes the hepatitis B surface antigen (HBsAg) to deliver CS epitopes, an approach developed by SmithKline Beecham (SKB) that is disclosed in U.S. Patent No. 5,928,902 that issued on July 27, 1999. That patent *inter alia* discloses a hybrid protein comprised of all of the C-terminal portion of the CS protein, four or more tandem repeats of the CS immunodominant region and the hepatitis B surface antigen. The CS epitopes include the

NANP repeat, in concert with additional CS epitopes,
including the T cell site identified by Berzofsky and Good
[Good et al., *Science*, (1987) 235(4792): p. 1059-62] (but
not the universal form developed by Nardin and co-worker
[Moreno et al., *Int. Immunol.*, (1991) 3(10): p. 997-1003
and Calvo-Calle et al., *J. Immunol.*, (1997) 159(3): p.
1362-1373]), fused to the hepatitis B surface protein.

This vaccine was recently the subject of human
clinical trials [Stoute et al., *N. Engl. J. Med.*, [1997]
336(2): p. 86-91]. When administered with one of three
different adjuvants, this vaccine protected 1/7, 2/7 and
6/7 individuals, respectively. Of the seven individuals
immunized with vaccine 2 (adjuvant: oil-in-water
emulsion), none of the five patients with anti-CS titers
(IFA) in the range of 100-12,800 were protected, whereas
the two vaccine recipients with antibody titers in the
range of 25,600-51,200 were both protected. Again,
protection was correlated positively with anti-CS titers.
It is interesting to note that the one patient that
received vaccine 1, the alum/oil-in-water formulation,
remained protected for at least six months.

The preliminary efficacy report for SKB's
malaria vaccine candidate (RTS,S) [Stoute et al.
(1997) *N. Engl. J. Med.*, 336(2): p. 86-91] although
encouraging, was tempered by the lack of long-term
protection in follow-up studies [Stoute et al. (1998
Oct) *J.infect Dis.*, 178(4):1139-44]. It was also
apparent that the use of a potent and complex
adjuvant (SBAS2) containing the immunostimulants QS-
21 and monophosphoryl lipid A (MPL), formulated in an
oil-in-water emulsion, was essential to achieve
efficacy, because volunteers receiving the vaccine
formulated on alum were not protected. Five of six
patients, who were initially protected after

administration of the RTS,S/SBAS2 formulation, were
not protected six months after receiving the third
vaccine dose. Similar results were recently reported
at the 48th annual meeting of the American Society of
Tropical Medicine and Hygiene for a field trial
conducted in Africa.

Like HBsAg, HBcAg is a particulate protein
derived from the hepatitis B virus that has been proposed
as a carrier for heterologous epitopes. The relative
immunogenicity of HBsAg (HBs) has been compared with HbcAg
(HBc), and the ability of each to evoke immune responses
in different genetic backgrounds [Milich et al., *Science*,
(1986) **234**(4782): p. 1398-1401]. These data emphasize the
higher immunogenicity of HBc relative to HBs, and the
universal responsiveness to HBc, irrespective of genetic
background.

For example, HBc is more than 300 times more
immunogenic than HBs in BALB/c mice; and, although both
B10.S and B10.M mice are non-responders to HBs, every
strain tested is responsive to HBc. These results re-
emphasize the suitability of HBc as a vaccine carrier and
specifically, its superiority over HBs, hence the
selection of HBc as opposed to HBs to carry heterologous
epitopes. These facets of HBc are thought to be
particularly important in malaria vaccine development,
because they address the genetic restriction and
inadequate antibody titers that have been largely
responsible for the inability to develop an effective
vaccine using the neutralizing CS epitopes.

The positive correlation between protection
against malaria infection and anti-CS antibody titer has
been demonstrated repeatedly over the past 15 years
[Etlinger et al., *Immunology*, (1988) 64(3): p. 551-558;
Etlinger et al., *J. Immunol.*, (1988) 140(2): p. 626-633;

Ballou et al., *Lancet*, (1987) 1(8545): p. 1277-1281;
Stoute et al., *N. Engl. J. Med.*, (1997) 336(2): p. 86-91
and Herrington *et al.*,. *Am J Trop Med Hyg*, (1991) 45(6):
p. 695-701]. The evidence that a vaccine eliciting high-
titer, long-lived antibody responses in sufficient vaccine
recipients can be protective suggests that protection
against malaria infection is achievable via anti-
sporozoite antibody production.

Using rodent models of malaria, it has been
found that malaria CS-repeats fused to the immunodominant
loop of HBc were able to protect mice against both *P.
berghei* and, perhaps more impressively, *P. yoelii* to
levels of 90-100 percent [Schodel et al., *Behring Inst.
Mitt.*, 1997(98): p. 114-119 and Schodel et al., *J. Exp.
Med.*, (1994) 180(3): p. 1037-46]. Further, antibody
responses to the *P. berghei* particle were shown to prime
antibody responses effectively over a wide range of
genetic backgrounds, confirming the universal priming
effects of HBc [Schodel et al., *J. Exp. Med.*, (1994)
180(3): p. 1037-46].

Another advantage of the HBc carrier is the
fact that it does not require complex adjuvants for
efficacy. This is due to the high inherent
immunogenicity of the particle. A comparison of the
immunogenicity of HBc-*P. berghei* particles showed
that alum, which is approved for human use, was more
effective than either IFA or CFA [Schodel et al., *J.
Exp. Med.*, (1994) 180(3): p. 1037-46]. The
importance of this observation is highlighted by
toxicity problems associated with newer, more complex
adjuvants as was recently noted in clinical trials of
SKB's candidate malaria vaccine [Stoute et al., *N.
Engl. J. Med.*, [1997] 336(2): p. 86-91].

The immunodominant B cell epitope of the CS protein of *P. falciparum*, which has been more widely studied than *P. vivax*, is a highly conserved repeated tetrapeptide (NANP) [Zavala *et al.*, *Science*, (1985) 228(4706): p. 1436-40], and antibodies to this epitope have been shown to be sporozoite-neutralizing in protecting against rodent and human malaria. Immune responsiveness to this epitope has been positively correlated with immunity to malaria in both vaccine recipients and naturally infected individuals. Indeed, a review of clinical trials data for pre-erythrocytic vaccines described previously (HBs-CS, FSV-1, NANP-TT), highlights a strong correlation between antibody titer and protection. Those individuals who have been protected by previous vaccine candidates have been associated with the highest anti-NANP antibody titers, with the possible exception of SKB's candidate vaccine (# 3 - RTS,S and adjuvant SBAS2 containing MPL and QS-21 in a water-in-oil formulation) where adjuvants appeared to play a critical role in protection, because protection was not long-lived, as noted before.

The family hepadnaviridae are enveloped DNA-containing animal viruses that can cause hepatitis B in humans (HBV). The hepadnavirus family includes hepatitis B viruses of other mammals, e.g., woodchuck (WHV), and ground squirrel (GSHV), and avian viruses found in ducks (DHV) and herons (HeHV). Hepatitis B virus (HBV) used herein refers to a member of the family hepadnaviridae, unless the discussion is referring to a specific example.

The nucleocapsid or core of the mammalian hepatitis B virus (HBV or hepadnavirus) contains a

sequence of 183 or 185 amino acid residues, depending on viral subtype, whereas the duck virus capsid contains 262 amino acid residues. Hepatitis B core protein monomers self-assemble into stable aggregates known as hepatitis B core protein particles (HBc particles). Two three-dimensional structures are reported for HBc particles. A first that comprises a minor population contains 90 copies of the HBc subunit protein as dimers or 180 individual monomeric proteins, and a second, major population that contains 120 copies of the HBc subunit protein as dimers or 240 individual monomeric proteins. These particles are referred to as T = 4 or T = 3 particles, respectively, wherein "T" is the triangulation number. These human HBc particles are about are about 30 or 34 nm in diameter, respectively. Pumpens et al., (1995) *Intervirology*, 38:63-74; and Metzger et al., (1998) *J. Gen. Viol.*, 79:587-590. See also, Wynne et al., (June 1999) *Mol. Cell*, 3:771-780.

Conway et al., (1997) *Nature*, 386:91-94, describe the structure of human HBc particles at 9 Ångstrom resolution, as determined from cryo-electron micrographs. Bottcher et al. (1997), *Nature*, 386:88-91, describe the polypeptide folding for the human HBc monomers, and provide an approximate numbering scheme for the amino acid residues at which alpha helical regions and their linking loop regions form. Zheng et al. (1992), *J. Biol. Chem.*, 267(13):9422-9429 report that core particle formation is not dependent upon the arginine-rich C-terminal domain, the binding of nucleic acids or the formation of disulfide bonds based on their study of mutant proteins lacking one or more cysteines and others'

work with C-terminal-truncated proteins [Birnbaum et al., (1990) *J.Virol*. 64, 3319-3330].

The nucleocapsid or viral core protein (HBc) has been disclosed as an immunogenic carrier moiety that stimulates the T cell response of an immunized host animal. See, for example, U.S. Patents No. 4,818,527, No 4,882,145 and No. 5,143,726. A particularly useful application of this carrier is its ability to present foreign or heterologous B cell epitopes at the site of the immunodominant loop that is present at about residue positions 70-90, and more usually recited as about positions 75 through 85 from the amino-terminus (N-terminus) of the protein. Clarke et al. (1991) F. Brown et al. eds., *Vaccines 91*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp.313-318.

During viral replication, HBV nucleocapsids associate with the viral RNA pre-genome, the viral reverse transcriptase (Pol), and the terminal protein (derived from Pol) to form replication competent cores. The association between the nucleocapsid and the viral RNA pre-genome is mediated via an arginine-rich domain at the carboxyl-terminus (C-terminus). When expressed in heterologous expression systems, such as *E.coli* where viral RNA pre-genome is absent, the protamine-like C-terminus; i.e., residues at positions 150 through 183, binds *E.coli* RNA.

In an application as a vaccine carrier moiety, it is preferable that the HBV nucleocapsids not bind nucleic acid derived from the host. Birnbaum et al., (1990) *J.Virol*. 64, 3319-3330 showed that the protamine-like C-terminal domain of HBV nucleocapsids could be deleted without interfering with the protein's ability to assemble into virus-

-14-

like particles. It is thus reported that proteins truncated to about position 144; i.e., containing the HBc sequence from position one through about 144, can self-assemble, whereas deletions beyond residue 139 abrogate capsid assembly [Seifer et al., (1995) Intervirology, 38:47-62].

More recently, Metzger et al., (1998) *J. Gen. Viol.*, 79:587-590 reported that the proline at position 138 (Pro-138 or P138) of the human sequence is required for particle formation. Those authors also reported that assembly capability of particles truncated at the carboxy-terminus to lengths of 142 and 140 residues was affected, with assembly capability being completely lost with truncations resulting in lengths of 139 and 137 residues.

Several groups have shown that truncated particles exhibit reduced stability relative to standard hepatitis B core particles [Gallina et al. (1989) *J.Virol.*, 63:4645-4652; Inada, et al. (1989) *Virus Res.*, 14:27-48], evident by variability in particle sizes and the presence of particle fragments in purified preparations [Maassen et al., (1994) *Arch. Virol.*, 135:131-142]. Thus, prior to the report of Metzger et al., above, Pumpens et al., (1995) *Intervirology*, 38:63-74 summarized the literature reports by stating that the carboxy-terminal border for HBc sequences required for self-assembly was located between amino acid residues 139 and 144, and that the first two or three amino-terminal residues could be replaced by other sequences, but elimination of four or eleven amino-terminal residues resulted in the complete disappearance of chimeric protein in transformed *E. coli* cells.

Recombinantly-produced hybrid particles bearing internal insertions (referred to in the art as HBc chimeric particles or HBc chimers) often appear to have a less ordered structure, when analyzed by electron microscopy, compared to particles that lack heterologous epitopes [Schodel et al., (1994) *J.Exp.Med.*, 180:1037-1046]. In some cases the insertion of heterologous epitopes into C-terminally truncated HBc particles has such a dramatic destabilizing affect that hybrid particles cannot be recovered following heterologous expression [Schodel et al. (1994) *Infect. Immunol.*, 62:1669-1676]. Thus, many chimeric HBc particles are so unstable that they fall apart during purification to such an extent that they are unrecoverable or they show very poor stability characteristics, making them problematic for vaccine development.

Chimeric hepatitis B core particles have been prepared by heterologous expression in a wide variety of organisms, including *E.coli*, *B.subtilis*, *Vaccinia*, *Salmonella typhimurium*, *Saccharomyces cerevisiae*. See, for example Pumpens et al., (1995) *Intervirology*, 38:63-74 , and the citations therein that note the work of several research groups, other than the present inventors.

A structural feature whereby the stability of full-length HBc particles could be retained, while abrogating the nucleic acid binding ability of full-length HBc particles, would be highly beneficial in vaccine development using the hepadnaviral nucleocapsid delivery system. Indeed, Ulrich et al. in their recent review of the use of HBc chimers as carriers for foreign epitopes [*Adv. Virus Res.*, vol.50 (1998) Academic Press pages 141-182] note

three potential problems to be solved for use of
those chimers in human vaccines. A first potential
problem is the inadvertent transfer of nucleic acids
in a chimer vaccine to an immunized host. A second
potential problem is interference from preexisting
immunity to HBc. A third possible problem relates to
the requirement of reproducible preparation of intact
chimer particles that can also withstand long-term
storage.

Initial evaluation of a particle displaying
epitopes from *P. falciparum* [CS-2; Schodel et al., *J. Exp.
Med.*, (1994) **180**(3): p. 1037-46] was encouraging.
However, using that particle as an immunogen in a vaccine
in mice, provided antibody titers that were lower than
those observed for the *P. berghei* and *P. yoelii* particles.

There are recognized to be two main CS-repeat
epitopes associated with *P. vivax* (type-I and type-II),
and a third, reported in 1993 and called 'vivax-like',
which is identical to the CS-repeat from the monkey
parasite (*P. siminovale*) resembling *P. ovale* [Qari et al.,
Lancet, 1993. **341**(8848): p. 780-783]. For simplicity,
this CS-repeat is referred to herein as *P. vivax* type-III.

The benefits of the inclusion of a universal T (Th)
cell epitope derived from the malaria parasite are
several-fold. First, the priming of malaria-specific Th
cells ensures that, should a vaccine recipient be exposed
to malaria, a more rapid and stronger anti-malaria
response is activated due to previous priming of malaria
specific T-helper cells. Secondly, vaccinees living in
malaria endemic regions experience natural 'boosting'
every time they are exposed to the parasite, because their
immune systems have been primed at both the B and Th cell
level. This effect is similar to clinical boosting by re-

vaccination, a process that can be difficult to coordinate in developing countries where malaria is endemic.

Although the CS gene is largely invariant, limited sequence variation has been noted to occur mainly in the immunodominant T cell epitopic domains. The fact that genetic mutations always appear to result in amino acid substitutions suggests that pressure at the protein level, possibly immunological pressure, has selected for variation. Typically, the problems associated with amino acid variability of an epitope can only be resolved by the inclusion of multiple variants of the epitope. However, Nardin and coworkers at New York University recently identified a consensus form of the T cell epitope CS 326-345 that appears to bind all class II MHC molecules [Calvo-Calle et al., *J. Immunol.*, (1993) **150**(4): p. 1403-1412 and Moreno et al., *J. Immunol.*, (1993) **151**(1): p. 489-499].

Studies have shown that this consensus epitope is 'universal', like the T cell help afforded by HBc, and suggests that it primes malaria-specific Th cells in essentially all vaccine recipients. The fact that this epitope of the CS protein was identified by CD4[+] T cells of volunteers that were protected against malaria following exposure to irradiated sporozoites, confirms that it is efficiently processed and presented *in vivo* by antigen presenting cells (APC) when presented in the context of sporozoite [Moreno et al., *Int. Immunol.*, (1991) **3**(10): p. 997-1003]. The identification of this epitope was a significant advancement in the task of developing a pre-erythrocytic stage malaria vaccine.

As disclosed hereinafter, the present invention provides a contemplated HBc chimer that provides unexpectedly high titers of antibodies against malaria sporozoites, and in one aspect also

provides a solution to the problems of HBc chimer
stability as well as the substantial absence of
nucleic acid binding ability of the construct. In
addition, a contemplated recombinant chimer exhibits
minimal, if any, antigenicity toward preexisting
anti-HBc antibodies.

BRIEF SUMMARY OF THE INVENTION

The present invention contemplates an
immunogen for inducing antibodies to the malaria-
causing parasite, *Plasmodium*, and particularly the
species P. falciparum and *P. vivax*, and a vaccine
comprising that immunogen dispersed in a
physiologically tolerable diluent. A contemplated
immunogen is a recombinant hepatitis B virus core
(HBc) protein chimer molecule with a length of about
140 to about 310 amino acid residues that contains
four peptide-linked amino acid residue sequence
domains from the N-terminus that are denominated
Domains I, II, III and IV.

The first domain, Domain I, comprises about
71 to about 85 amino acid residues whose sequence
includes at least the sequence of the residues of
position 5 through position 75 of HBc.

The second domain, Domain II, comprises
about 18 to about 58 amino acid residues peptide-
bonded to residue 75 of which (i) a sequence of HBc
is present from HBc positions 76 through 85 and (ii)
a sequence of 8 to about 48 residues that constitute
a B cell epitope of the CS protein of a species of
the parasite *Plasmodium* that is peptide-bonded
between the HBc residues of positions 78 and 79.

The third domain, Domain III, is an HBc
sequence from position 86 through position 135
peptide-bonded to residue 85.

The fourth domain, Domain IV, comprises (i)
zero to fourteen residues of a HBc amino acid residue
sequence from position 136 through 149 peptide-bonded
to the residue of position 135 of Domain III, (ii)
zero to three cysteine residues, (iii) fewer than
three arginine or lysine residues, or mixtures
thereof adjacent to each other, and (iv) up to about
100 amino acid residues in a sequence heterologous to
HBc from position 150 to the C-terminus, with the
proviso that at least five amino acid residues are
present of the amino acid residue sequence from
position 136 through 149; i. e., residues of
positions 136-140, when (a) zero cysteine residues
are present and (b) fewer than about five
heterologous amino acid residues are present, thus,
Domain IV contains at least 5 residues.

In preferred embodiments, the immunogen is
in the form of self-assembled particles and the
Plasmodium B cell epitope is that of *P. falciparum* or
P. vivax. It is also preferred that the HBc sequence
of Domain I includes the residues of position 1
through position 75 with no additional residues at
the N-terminus. It is further preferred that a
contemplated immunogen contain one cysteine residue
within Domain IV in an amino acid residue sequence
heterologous to that of HBc from position 150 to the
C-terminus. It is particularly preferred that that
heterologous sequence comprise a T cell epitope from
the same species of *Plasmodium* as the B cell epitope.

Another embodiment comprises an inoculum or
vaccine that comprises an above HBc chimer particle

that is dissolved or dispersed in a pharmaceutically
acceptable diluent composition that typically also
contains water. When administered in an immunogenic
effective amount to an animal such as a mammal or
bird, an inoculum induces antibodies that immunoreact
specifically with the chimer particle or the
conjugated (pendently-linked) hapten. The antibodies
so induced also preferably immunoreact specifically
with (bind to) an antigen containing the hapten, such
as a protein where the hapten is a peptide or a
saccharide where the hapten is an oligosaccharide.

The present invention has several benefits
and advantages.

A particular benefit of the invention is
that its use as a vaccine provides extraordinary
antibody titers against the *Plasmodium* species of the
B cell epitope.

An advantage of the invention is that those
very high antibody titers have been produced with the
aid of an adjuvant approved for use in humans.

Another benefit of the invention is that
the recombinant immunogen is prepared easily and
using well known cell culture techniques.

Another advantage of the invention is that
the immunogen is easily prepared using well known
recombinant techniques.

Yet another benefit of the invention is
that a preferred immunogen exhibits greater stability
at elevated temperatures than to other HBc chimers.

Yet another advantage of the invention is
that a contemplated immunogen is substantially free
of nucleic acids.

Still further benefits and advantages will
be apparent to the worker of ordinary skill from the
disclosure that follows.

BRIEF DESCRIPTION OF THE DRAWINGS

In the drawings forming a portion of this
disclosure:

Fig. 1 shows the modifications made to
commercial plasmid vector pKK223-3 in the preparation
of plasmid vector pKK223-3N used herein for
preparation of recombinant HBc chimers. The modified
sequence (SEQ ID NO:180) is shown below the sequence
of the commercially available vector (SEQ ID NO:181).
The bases of the added NcoI site are shown in lower
case letters and the added bases are shown with
double underlines, whereas the deleted bases are
shown as dashes. The two restriction sites present
in this segment of the sequence (NcoI and HindIII)
are indicated.

Fig. 2, shown in three panels as Figs. 2A,
2B and 2C, schematically illustrates a preferred
cloning strategy in which a malarial B cell epitope
such as (NANP)$_4$ (SEQ ID NO:1) is cloned into the
EcoRI and SacI sites of an engineered HBc gene (Fig.
2A) between positions 78 and 79, which destroys the
EcoRI site, while preserving the SacI site. Fig. 2B
shows DNA (double stranded and encoded amino acid
residue sequence) that encodes a T cell epitope such
as that referred to as Pf/CS326-345 (Pf/CS-UTC) and a
stop codon (SEQ ID NOs:79 and 80) cloned into the
EcoRI and HindIII sites at the C-terminus of an
engineered, truncated HBc gene containing the first
149 HBc residues (HBc149). PCR amplification of the
construct of Fig. 2B using a primer having a 5'-

terminal SacI restriction site adjacent to a HBc-
encoding sequence beginning at residue position 79
followed by digestion of the amplified sequence and
the construct of Fig. 2A with SacI, followed by
ligation of the appropriate portions is shown in Fig.
2C to form a single gene construct referred to
hereinafter as V12 that encodes a B cell- and T cell-
containing immunogen for a vaccine against *P.
falciparum*.

　　　　Fig. 3 is a photograph of an SDS-PAGE
analysis under reducing conditions to show the
stabilizing effects on expressed particles of a codon
for a single cysteine residue inserted in frame
between the C-terminal codon (V149) and the
termination codon of HBc in a chimer that also
contains (NANP)$_4$ inserted between the amino acids of
positions 78 and 79 (V2.Pf1+C), and a similar
construct whose C-terminus is residue V149 (V2.Pf1)
at day zero and after 15 days at 37°C. [Lane 1,
V2.Pf1 - day 0; Lane 2, V2.Pf1 - day 15 at 37°C; Lane
3, V2.Pf1+C, day 0; Lane 4, V2.Pf1+C - day 15 at
37°C.]

　　　　Fig. 4 is a photograph of an SDS-PAGE
analysis under reducing conditions that illustrates
the stabilizing effects on chimer HBc149 particles
containing (NANP)$_4$ inserted between amino acids 78
and 79 and the cysteine-containing T cell epitope
fused to the C-terminus [V2.Pf1+Pf/CS-UTC also
referred to as V12.Pf1] as compared to a similar
particle in which the C-terminal Cys was replaced by
an Ala residue [V2.Pf1+ Pf/CS-UTC(C17A) also referred
to as V12.Pf1(C17A)] at day zero and after 28 days at
37°C. [Lane 1, V2.Pf1+Pf/CS-UTC - day zero; Lane 2,

V2.Pf1+ Pf/CS-UTC - day 28 at 37°C; Lane 3,
V2.Pf1+Pf/CS-UTC(C17A) - day zero; Lane 4, V2.Pf1+
Pf/CS-UTC(C17A) - day 28 at 37°C.]

Fig. 5 is a graph showing the results of an
indirect immunofluorescence assay (IFA) carried out
using glutaraldehyde-fixed *P. falciparum* sporozoites
and FITC-labeled anti-mouse IgG (gamma-chain
specific) to detect bound antibody titers (log of
1/dilution; ordinate) over time in weeks (abscissa)
for three chimeric immunogens after immunization in
mice. Data for the prior art chimer immunogen, CS-2,
are shown as squares, those for the recombinant HBc
chimer V12.Pf1 are shown as diamonds, whereas those
for the recombinant HBc chimer V12.Pf3.1 are shown as
triangles.

Fig.6, shown in two panels as Fig. 6A and
Fig. 6B, provides an alignment of six published
sequences for mammalian HBc proteins from six
viruses. The first (SEQ ID NO:170), human viral
sequence is of the *ayw* subtype and was published in
Galibert et al. (1983) *Nature*, **281**:646-650; the
second human viral sequence (SEQ ID NO:171), of the
adw subtype, was published by Ono et al. (1983)
Nucleic Acids Res., **11(6)**: 1747-1757; the third human
viral sequence (SEQ ID NO:172), is of the *adw2*
subtype and was published by Valenzuela et al.,
Animal Virus Genetics, Field et al. eds., Academic
Press, New York (1980)pages 57-70; the fourth human
viral sequence (SEQ ID NO:173), is of the *adyw*
subtype that was published by Pasek et al. (1979)
Nature, **282**:575-579; the fifth sequence (SEQ ID
NO:174), is that of the woodchuck virus that was
published by Galibert et al. (1982) *J. Virol.*, **41**:51-
65; and the sixth mammalian sequence, (SEQ ID

NO:168), is that of the ground squirrel that was
published by Seeger et al. (1984) *J. Virol.*,51:367-
375.

Figure 7 is a photograph of an SDS-PAGE
analysis under reducing conditions following
incubations at 37°C for 0, 1 and 2 days that
illustrates the stabilizing effects on (1) chimer
HBc149 particles containing the *P. falciparum* (NANP)$_4$
immunogenic sequence inserted between HBc amino acid
residues 78 and 79 that also contain a carboxy-
terminal universal *P. falciparum* malarial T cell
epitope peptide-bonded to HBc position 149 [UTC;
V12.Pf1 = V2.Pf1 + Pf/CS-UTC], and (2) similar
particles in which the cysteine at position 17 of the
UTC was mutated to be an alanine residue and a
cysteine residue was added at residue position 150,
between the HBc residue at position 149 and the
beginning of the UTC [V12.Pf1(C17A)+C150].
Definitions

Numerals utilized in conjunction with HBc
chimers indicate the position in the HBc *ayw* amino
acid residue sequence of SEQ ID NO:170 at which one
or more residues has been added to the sequence,
regardless of whether additions or deletions to the
amino acid residue sequence are present. Thus,
HBc149 indicates that the chimer ends at residue 149,
whereas HBc149 + C150 indicates that that same chimer
contains a cysteine residue at HBc position 150. On
the other hand, the malarial CS protein universal T
cell epitope (UTC) is 20 residues long, and a
replacement of the cysteine at position 17 in that
sequence by an alanine is referred to as CS-
UTC(C17A).

The term "antibody" refers to a molecule that is a member of a family of glycosylated proteins called immunoglobulins, which can specifically bind to an antigen.

The word "antigen" has been used historically to designate an entity that is bound by an antibody or receptor, and also to designate the entity that induces the production of the antibody. More current usage limits the meaning of antigen to that entity bound by an antibody or receptor, whereas the word "immunogen" is used for the entity that induces antibody production or binds to the receptor. Where an entity discussed herein is both immunogenic and antigenic, reference to it as either an immunogen or antigen is typically made according to its intended utility.

"Antigenic determinant" refers to the actual structural portion of the antigen that is immunologically bound by an antibody combining site or T-cell receptor. The term is also used interchangeably with "epitope".

The word "conjugate" as used herein refers to a hapten operatively linked to a carrier protein, as through an amino acid residue side chain.

The term "conservative substitution" as used herein denotes that one amino acid residue has been replaced by another, biologically similar residue. Examples of conservative substitutions include the substitution of one hydrophobic residue such as isoleucine, valine, leucine or methionine for another, or the substitution of one polar residue for another such as between arginine and lysine, between glutamic and aspartic acids or between glutamine and asparagine and the like.

The term "corresponds" in its various grammatical forms as used in relation to peptide sequences means the peptide sequence described plus or minus up to three amino acid residues at either or both of the amino- and carboxy-termini and containing only conservative substitutions in particular amino acid residues along the polypeptide sequence.

The term "Domain" is used herein to mean a portion of a recombinant HBc chimer molecule that is identified by (i) residue position numbering relative to the position numbers of HBcAg subtype *ayw* as reported by Galibert et al., (1979) *Nature*, 281:646-650 (SEQ ID NO:170). The polypeptide portions of at least chimer Domains I, II and III are believed to exist in a similar tertiary form to the corresponding sequences of naturally occurring HBcAg.

As used herein, the term "fusion protein" designates a polypeptide that contains at least two amino acid residue sequences not normally found linked together in nature that are operatively linked together end-to-end (head-to-tail) by a peptide bond between their respective carboxy- and amino-terminal amino acid residues. The fusion proteins of the present invention are HBc chimers that induce the production of antibodies that immunoreact with a polypeptide or pathogen-related immunogen that corresponds in amino acid residue sequence to the polypeptide or pathogen-related portion of the fusion protein.

The phrase "hepatitis B" as used here refers in its broadest context to any member of the family hepadnaviridae, as discussed before.

The term "residue" is used interchangeably with the phrase amino acid residue.

DETAILED DESCRIPTION OF THE INVENTION

The present invention contemplates an
immunogen and a vaccine comprising that immunogen
against the malaria parasite, particularly those that
infect humans; i.e., *P. falciparum* and *P. vivax.*
Historically, one of the main shortfalls of peptide-
based vaccines has been the lack of persistence of
antibody following immunization. As discussed
hereinafter, using the HBc chimer immunogen applied
to *P. falciparum* vaccine development, high titers of
neutralizing antibody are maintained for more than 6
months in mice following a 2-dose immunization
regimen. This is consistent with and superior to the
protection studies in the *P. yoelii* model using a
similar but differently constructed immunogen, in
which immunity obtained from challenge infection was
evident 3 months after immunization using an
immunogen different from that used here. [Schodel et
al., *Behring Inst. Mitt.*, 1997(98): p. 114-119.]

A contemplated immunogen is a recombinant
hepatitis B virus core (HBc) protein chimer molecule
with a length of about 140 to about 310 and
preferably about 155 to 235 amino acid residues that
contains four peptide-linked amino acid residue
sequence domains from the N-terminus that are
denominated Domains I, II, III and IV.

(a) Domain I comprises about 71 to about 85
amino acid residues whose sequence includes at least
the sequence of the residues of position 5 through
position 75 of HBc.

(b) Domain II comprises about 11 to about
58 amino acid residues peptide-bonded to residue 75.
This sequence includes (i) a sequence of HBc from HBc

positions 76 through 85 and (ii) a sequence of 8 to about 48 residues that constitute a B cell epitope of the circumsporozoite protein of a species of the parasite *Plasmodium* that is peptide-bonded between the HBc residues of positions 78 and 79.

(c) Domain III is an HBc sequence from position 86 through position 135 that is peptide-bonded to residue 85.

d) Domain IV comprises (i) zero to fourteen residues of a HBc amino acid residue sequence from position 136 through 149 peptide-bonded to the residue of position 135 of Domain III, (ii) zero to three cysteine residues, (iii) fewer than three arginine or lysine residues, or mixtures thereof adjacent to each other, and (iv) up to 100, and preferably up to 25, amino acid residues in a sequence heterologous to HBc from position 150 to the C-terminus, with the proviso that at least five amino acid residues of the amino acid residue sequence from position 136 through 149; i.e., residues of positions 136-140, are present when (a) zero cysteine residues are present and (b) fewer than about five heterologous amino acid residues are present.

In examining the length of a contemplated HBc chimer, such a recombinant protein can have a length of about 140 to about 310 amino acid residues. Preferably, that length is about 155 to about 235 residues. More preferably, the length is about 165 to about 210 residues. Most preferably, the length is about 190 to about 200 residues. These differences in length arise from changes in the length of Domains I, II and IV.

HBc chimers having a Domain I that contains more than a deletion of the first three amino-

terminal (N-terminal) residues have been reported to result in the complete disappearance of HBc chimer protein in *E. coli* cells. Pumpens et al.,(1995) *Intervirology*, 38:63-74. On the other hand, a recent study in which an immunogenic 23-mer polypeptide from the influenza M2 protein was fused to the HBc N-terminal sequence reported that the resultant fusion protein formed particles when residues 1-4 of the native HBc sequence were replaced. Neirynck et al. (October 1999) *Nature Med.*, 5(10):1157-1163. Thus, the art teaches that particles can form when an added amino acid sequence is present peptide-bonded the one of residues 1-5 of HBc, whereas particles do not form if no additional sequence is present and more than residues 1-3 are deleted from the N-terminus of HBc.

An N-terminal sequence peptide-bonded to one of the first five N-terminal residues of HBc can contain a sequence of up to about 25 residues that are heterologous to HBc. Exemplary sequences include a B cell or T cell epitope such as those discussed hereinafter, a sequence of another (heterologous) protein such as β-galactosidase as can occur in fusion proteins as a result of the expression system used, or another hepatitis B-related sequence such as that from the Pre-S1 or Pre-S2 regions or the major HbsAg immunogenic sequence.

Domain I preferably has the sequence of residues of positions 1 through 75 of HBc, and is free of added residues at the amino-terminus (N-terminus). Domain I is also therefore preferably free of deletions of residues of positions 1-3.

Domain II, which is peptide-bonded to residue 75, contains the sequence of HBc residues of positions 76 through 85, and has a malarial B cell

epitope whose length is 8 through about 28 residues peptide-bonded between residues 78 and 79. Preferred malarial B cell epitopes are discussed hereinafter.

Preferred malarial B cell epitopes for insertion between residues 78 and 79 of a recombinant HBc chimer are enumerated in Table A, below.

Table A
Malarial B Cell Epitopes

P. falciparum

$(NANP)_4$	SEQ ID NO:1
$NANPNVDP(NANP)_3NVDP$	SEQ ID NO:2
$NANPNVDP(NANP)_3$	SEQ ID NO:3
$(NANP)_3NVDPNANP$	SEQ ID NO:4
$NANPNVDP(NANP)_3NVDPNANP$	SEQ ID NO:5
$NPNVDP(NANP)_3NV$	SEQ ID NO:6
$NPNVDP(NANP)_3NVDP$	SEQ ID NO:7
$NPNVDP(NANP)_3NVDPNA$	SEQ ID NO:8
$NVDP(NANP)_3NV$	SEQ ID NO:9
$NVDP(NANP)_3NVDP$	SEQ ID NO:10
$NVDP(NANP)_3NVDPNA$	SEQ ID NO:11
$DP(NANP)_3NV$	SEQ ID NO:12
$DP(NANP)_3NVDP$	SEQ ID NO:13
$DP(NANP)_3NVDPNA$	SEQ ID NO:14

P. vivax

DRAAGQPAGDRADGQPAG	SEQ ID NO:15
ANGAGNQPGANGAGDQPGA- NGADNQPGANGADDQPG	SEQ ID NO:16
ANGAGNQPGANGAGDQPG	SEQ ID NO:17
ANGADNQPGANGADDQPG	SEQ ID NO:18

```
ANGAGNQPGANGADNQPG          .            SEQ ID NO:19
ANGADNQPGANGADDQPG                       SEQ ID NO:20
APGANQEGGAAAPGANQEGGAA                   SEQ ID NO:21
```

P. bergeii

(DP$_4$NPN)$_2$ SEQ ID NO:22

P. yoelli

(QGPGAP)$_4$ SEQ ID NO:23

Domain III contains the sequence of HBc position 86 through position 135 peptide-bonded at its N-terminus to residue 85.

Domain IV comprises (i) zero to fourteen residues of a HBc amino acid residue sequence from position 136 through 149 peptide-bonded to the residue of position 135 of Domain III, (ii) zero to three cysteine residues, (iii) fewer than three arginine or lysine residues, or mixtures thereof adjacent to each other, and (iv) up to about 100 amino acid residues in a sequence heterologous to HBc from position 150 to the C-terminus (typically as one or more T cell epitopes), with certain provisos. Although Domain IV can contain up to about 100 residues that are heterologous to HBc from position 150 through the C-terminus, this domain needs no residues in addition to those recited before to provide an effective immunogen.

However, when the chimeric protein ends at HBc residue 135, desired, particularly immunogenic particles do not form even when a C-terminal cysteine is present. On the other hand, desired particles do form when residues of positions 136-140 are present with or without an added C-terminal cysteine or when

(a) one cysteine residue is present and (b) about
five heterologous amino acid residues are also
present peptide-bonded to HBc residue 135. Put
differently, Domain IV can end at HBc residue 135 so
long as at least five heterologous residues are
present and a cysteine residue is also present.
Otherwise, Domain IV ends at least at HBc residue
140. Thus, Domain IV contains at least 5 amino acid
residues.

It is preferred that Domain IV contain up
to fourteen residues of an HBc sequence from position
136 through position 149 peptide-bonded to residue
135; i.e., an HBc sequence that begins with the
residue of position 136 that can continue through
position 149. Thus, if the residue of position 148
is present, so is the sequence of residues of
positions 136 through 147, or if residue 141 is
present, so is the sequence of residues of positions
136 through 140.

In one embodiment, Domain IV comprises a
sequence of HBc from residue 136 through 140 peptide-
bonded to the residue of position 135 of Domain III.
The remainder of Domain IV contains (i) zero to nine
residues of a HBc amino acid residue sequence from
position 141 through 149 peptide-bonded to the
position 136-140 sequence, (ii) zero to three
cysteine residues, (iii) fewer than three arginine or
lysine residues, or mixtures thereof adjacent to each
other, and (iv) up to 50 amino acid residues, and
more preferably up to about 25 residues, in a
sequence that constitutes a T cell epitope of the
same species of *Plasmodium* as the B cell epitope
peptide-bonded to the final HBc amino acid residue
present in the chimer or a cysteine residue. Thus,

for example, the T cell epitope can be bonded to the
carboxy-terminal-most HBc residue such as residue
149, or to a cysteine residue that is bonded to that
final HBc residue.

Domain IV can also contain zero to three
cysteine residues and those Cys residues are present
within about 30 residues of the carboxy-terminus (C-
terminus) of the chimer molecule. Preferably, one
cysteine (Cys) residue is present, and that Cys is
preferably present as the carboxy-terminal (C-
terminal) residue, unless a malarial T cell epitope
is present as part of Domain IV. When such a T cell
epitope is present, the preferred Cys is preferably
within the C-terminal last five residues of the HBc
chimer. Preferred malarial T cell epitopes are
discussed hereinafter.

The presence of the above-discussed
cysteine residue(s) provides an unexpected
enhancement of the ability of the chimer molecules to
form immunogenic particles, as well as unexpected
thermal stability to an immunogen particle (discussed
hereinafter). Thus, a preferred HBc chimer immunogen
tends to be stable to decomposition at 37°C to a
greater extent than does a similar chimer lacking
that cysteine residue. This enhanced stability is
illustrated in Figs. 3, 4 and 7, and is discussed
hereinafter.

Domain IV contains fewer than three
arginine or lysine residues, or mixtures thereof
adjacent to each other. Arginine and lysines are
present in the C-terminal region of HBc that extends
from position 150 through the C-terminus of the
native molecule. That region is sometimes referred
to in the are as the "protamine" or "arginine-rich"

region of the molecule and is thought to bind to nucleic acids. A contemplated HBc chimer molecule and particle are substantially free of bound nucleic acids.

The substantial freedom of nucleic acid binding can be readily determined by a comparison of the absorbance of the particles in aqueous solution measured at both 280 and 260 nm; i.e., a 280/260 absorbance ratio. The contemplated particles do not bind substantially to nucleic acids that are oligomeric and/or polymeric DNA and RNA species originally present in the cells of the organism used to express the protein. Such nucleic acids exhibit an absorbance at 260 nm and relatively less absorbance at 280 nm, whereas a protein such as a contemplated chimer absorbs relatively less at 260 nm and has a greater absorbance at 280 nm.

Thus, recombinantly expressed HBc particles or chimeric HBc particles that contain the arginine-rich sequence at residue positions 150-183 (or 150-185) exhibit a ratio of absorbance at 280 nm to absorbance at 260'nm (280:260 absorbance ratio) of about 0.8, whereas particles free of the arginine-rich nucleic acid binding region of naturally occurring HBc such as those that contain fewer than three arginine or lysine residues or mixtures thereof adjacent to each other, or those having a native or chimeric sequence that ends at about HBc residue position 140 to position 149, exhibit a 280:260 absorbance ratio of about 1.2 to about 1.6.

Chimeric HBc particles of the present invention are substantially free of nucleic acid binding and exhibit a 280:260 absorbance ratio of about 1.2 to about 1.6, and more typically, about 1.4

to about 1.6. This range is due in large part to the number of aromatic amino acid residues present in Domains II and IV of a given chimeric HBc particle. That range is also in part due to the presence of the Cys in Domain IV of a contemplated chimer, whose presence can diminish the observed ratio by about 0.1 for a reason that is presently unknown.

The contemplated chimer HBc particles are more stable in aqueous buffer at 37°C over a time period of about two weeks to about one month than are particles formed from a HBc chimer containing the same peptide-linked Domain I, II and III sequences and an otherwise same Domain IV sequence in which the one to three cysteine residues [C-terminal cysteine residue(s)] are absent or a single C-terminal residue present is replaced by another residue such as an alanine residue. Stability of various chimer particles is determined as discussed hereinafter.

Thus, for example, particles containing a heterologous malarial epitope in Domain II [e.g. $(NANP)_4$] and a single cysteine residue C-terminal to residue valine 149 is more stable than otherwise identical particles assembled from chimer molecules whose C-terminal residue is valine 149. Similarly, particles containing the above malarial B cell epitope in Domain II and the universal malarial T cell epitope that contains a single cysteine near the C-terminus are more stable than are otherwise identical particles in which that cysteine is replaced by an alanine residue. See, Figs. 3, 4 and 7 and the discussion relating thereto hereinafter.

A contemplated particle containing a C-terminal cysteine residue is also typically prepared in greater yield than is a particle assembled from a

chimer molecule lacking a C-terminal cysteine. This
increase in yield can be seen from the mass of
particles obtained or from integration of traces from
analytical gel filtration analysis using Superose® 6
HR as discussed hereinafter and shown in Tables 9A
and 9B.

Although the T cell help afforded by HBc is
highly effective in enhancing antibody responses
(i.e. B cell-mediated) to 'carried' epitopes
following vaccination, HBc does not activate malaria-
specific T cells, except in restricted individuals
for whom the B cell epitope is also a T cell epitope.
To help ensure universal priming of malaria-specific
T helper cells, in addition to B cells, one or more
malaria-specific T helper epitopes is preferably
incorporated into a contemplated immunogen and is
located in Domain IV of the immunogen.

A particularly preferred recombinant HBc chimer
includes a T cell epitope of the same _Plasmodium_ species
as the B cell epitope. Thus, where the B cell epitope of
Domain II is that of _P. falciparum_, the T cell epitope is
also that of _P. falciparum_, and the like. Using this
matching strategy, T cells are primed to the same species
as that to which antibodies are initially induced by the B
cell epitope. Particularly preferred T cell epitopes
present as a part of Domain IV are enumerated in Table B,
below.

Table B

Malarial Universal T Cell Epitope

P. falciparum

GIEYLNKIQNSLSTEWSPCSVT   SEQ ID NO:24

P. vivax

YLDKVRATVGTEWTPCSVT   SEQ ID NO:25

P. yoelli

EFVKQISSQLTEEWSQCSVT   SEQ ID NO:26

A plurality of the above or another T cell epitopes can be present in Domain IV or another B cell epitope can be present. In preferred practice, Domain IV has up to about 50 residues in a sequence heterologous to HBc. Most preferably, that sequence is up to about 25 residues and includes one of the universal T cell epitopes shown in Table B, above.

A contemplated recombinant HBc chimer molecule is typically present and is used in an immunogen or vaccine as a self-assembled particle. These particles are comprised of 180 to 240 chimer molecules that separate into protein molecules in the presence of disulfide reducing agents such as 2-mercaptoethanol, and the individual molecules are therefore thought to be bound together into the particle primarily by disulfide bonds. These particles are similar to the particles observed in patients infected with HBV, but these particles are non-infectious. Upon expression in various prokaryotic and eukaryotic hosts, the individual recombinant HBc chimer molecules assemble in the host

into particles that can be readily harvested from the host cells.

The amino acid sequence of HBc from residue position 1 through at least position 140 is preferably present in a contemplated chimer molecule and particle. The sequence from position 1 through position 149 is more preferably present. A malarial B cell epitope is present between residues 78 and 79 and a single cysteine residue or a malarial T cell epitope containing a cysteine residue is preferably present as a C-terminal addition to the HBc sequence as part of Domain IV. A contemplated recombinant HBc chimer is substantially free of bound nucleic acid. A preferred chimer particle that contains an added Cys residue at or near the C-terminus of the molecule is also more stable at 37°C than is a similar particle that does not contain that added Cys.

In addition to the before-discussed N- and C-truncations and insertion of malarial epitopes, a contemplated chimer molecule can also contain conservative substitutions in the amino acid residues that constitute HBc Domains I, II, III and IV. Conservative substitutions are as defined before.

More rarely, a "nonconservative" change, e.g., replacement of a glycine with a tryptophan is contemplated. Analogous minor variations can also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues can be substituted, inserted, or deleted without abolishing biological activity can be found using computer programs well known in the art, for example LASERGENE software (DNASTAR Inc., Madison, Wis.)

The HBc portion of a chimer molecule of the present invention [the portion having the HBc

sequence that has other than a sequence of an added
epitope, or heterologous residue(s) that are a
restriction enzyme artifact] most preferably has the
amino acid residue sequence at positions 1 through
149 of subtype *ayw* that is shown in Fig. 6 (SEQ ID
NO:170), when present. Somewhat less preferred are
the corresponding amino acid residue sequences of
subtypes *adw*, *adw2* and *adyw* that are also shown in
Fig. 6 (SEQ ID NOs:171, 172 and 173). Less preferred
still are the sequences of woodchuck and ground
squirrel at aligned positions 1 through 149 that are
the last two sequences of Fig 6 (SEQ ID NOs:174 and
168). As noted elsewhere, portions of different
sequences from different mammalian HBc proteins can
be used together in a single chimer.

        When the HBc portion of a chimer molecule
of the present invention has other than a sequence of
a mammalian HBc molecule at positions 1 through 149,
when present, because one or more conservative
substitutions has been made, it is preferred that no
more than 10 percent, and more preferably no more
than 5 percent, and most preferably no more than 3
percent of the amino acid residues are substituted as
compared to SEQ ID NO:170 from position 1 through
149. A contemplated chimer of 149 HBc residues can
therefore contain up to about 15 residues that are
different from those of SEQ ID NO:170 at positions 1
through 149, and preferably about 7 or 8 residues.
More preferably, up to about 5 residues are different
from the *ayw* sequence (SEQ ID NO:170) at residue
positions 1-149. Where a HBc sequence is truncated
further at one or both termini, the number of
substituted residues is proportionally different.
Deletions elsewhere in the molecule are considered

conservative substitutions for purposes of
calculation.

Chimer Preparation

A contemplated chimeric immunogen is
prepared using the well known techniques of
recombinant DNA technology. Thus, sequences of
nucleic acid that encode particular polypeptide
sequences are added and deleted from the precursor
sequence that encodes HBV.

As was noted previously, the HBc
immunodominant loop is usually recited as being
located at about positions 75 through 85 from the
amino-terminus (N-terminus) of the intact protein.
The malarial B cell epitope-containing sequence is
placed into that immunodominant loop sequence of
Domain II. That placement substantially eliminates
the HBc immunogenicity and antigenicity of the HBc
loop sequence, while presenting the malarial B cell
epitope in an extremely immunogenic position in the
assembled chimer particles.

One of two well-known strategies is
particularly useful for placing the malarial B cell
sequence into the loop sequence at the desired
location between residues 78 and 79. A first, less
successful strategy is referred to as replacement in
which DNA that codes for a portion of the loop is
excised and replaced with DNA that encodes a malarial
B cell sequence. The second strategy is referred to
as insertion in which a malarial B cell sequence is
inserted between adjacent residues in the loop.

Site-directed mutagenesis using the
polymerase chain reaction (PCR) is used in one
exemplary replacement approach to provide a chimeric

HBc DNA sequence that encodes a pair of different
restriction sites, e.g. EcoRI and SacI, one near each
end of the immunodominant loop-encoding DNA.
Exemplary residues replaced are 76 through 81. The
loop-encoding section is excised, a desired malarial
B cell epitope-encoding sequence flanked on each side
by appropriate HBc sequence residues is ligated into
the restriction sites and the resulting DNA is used
to express the HBc chimer. See, for example, Table 2
of Pumpens et al., (1995) *Intervirology*, **38**:63-74 for
exemplary uses of a similar technique.

Alternatively, a single restriction site or
two sites can be encoded into the region, the DNA cut
with a restriction enzyme(s) to provide "sticky" or
ends, and an appropriate sticky- or blunt-ended
heterologous DNA segment ligated into the cut region.
Examples of this type of sequence replacement into
HBc can be found in the work reported in Schodel et
al., (1991) F. Brown et al. eds., *Vaccines 91*, Cold
Spring Harbor Laboratory, Cold Spring Harbor, New
York, pp.319-325, Schodel et al., *Behring Inst.
Mitt.*, 1997(98): p. 114-119 and Schodel et al., *J.
Exp. Med.*, (1994) 180(3): p. 1037-4, the latter two
papers discussing the preparation of vaccines against
P. yoelii and *P. berghei*, respectively.

It has surprisingly been found that the
insertion position within the HBc immunogenic loop
and the presence of loop residues are of import to
the activity of the immunogen. Thus, as is
illustrated hereinafter, placement of a malarial B
cell epitope between HBc residue positions 78 and 79
provides a particulate immunogen that is ten to one
thousand times more immunogenic than placement of the
same immunogen in an excised and replaced region

between residues 76 and 81. In addition, placement
of the same malarial immunogen between residues 78
and 79 as compared to between residues 77 and 78
provided an unexpected enhancement of about 15-fold.
Thus, a replacement strategy that results in a net
removal of residues from the immunodominant loop is
not used herein.

Insertion is therefore preferred. In an
illustrative example of the insertion strategy, site-
directed mutagenesis is used to create two
restriction sites adjacent to each other and between
codons encoding adjacent amino acid residues, such as
those at residue positions 78 and 79. This technique
adds twelve base pairs that encode four amino acid
residues (two for each restriction site) between
formerly adjacent residues in the HBc loop.

Upon cleavage with the restriction enzymes,
ligation of the DNA coding for the malarial sequence
and expression of the DNA to form HBc chimers, the
HBc loop amino acid sequence is seen to be
interrupted on its N-terminal side by the two
residues encoded by the 5' restriction site, followed
toward the C-terminus by the malarial B-cell epitope
sequence, followed by two more heterologous, non-loop
residues encoded by the 3' restriction site and then
the rest of the loop sequence. This same strategy is
also preferably used for insertion into Domain IV of
a T cell epitope or one or more cysteine residues
that are not a part of a T cell epitope. A similar
strategy using an insertion between residues 82 and
83 is reported in Schoedel et al., (1990) F. Brown et
al. eds., *Vaccines 90*, Cold Spring Harbor Laboratory,
Cold Spring Harbor, New York, pp.193-198.

More specifically, this cloning strategy is
illustrated schematically in Figs. 2A, 2B and 2C. In
Fig. 2A, a DNA sequence that encodes a C-terminal
truncated HBc sequence (HBc149) is engineered to
contain adjacent EcoRI and SacI sites between
residues 78 and 79. Cleavage of that DNA with both
enzymes provides one fragment that encodes HBc
positions 1-78 3'-terminated with an EcoRI sticky
end, whereas the other fragment has a 5'-terminal
SacI sticky end and encodes residues of positions 79-
149. Ligation of a synthetic nucleic acid having a
5' AATT overhang followed by a sequence that encodes
a desired malarial B cell epitope and a AGCT
3'overhang provides a HBc chimer sequence that
encodes that B cell epitope flanked on each side by
two heterologous residues (GI and EL, respectively)
between residues 78 and 79, while destroying the
EcoRI site and preserving the SacI site.

A similar strategy is shown in Fig. 2B for
insertion of a cysteine-containing sequence, such as
a particularly preferred T cell epitope such as that
referred to asPF/CS326-345 (Pf-UTC). Here, EcoRI and
HindIII restriction sites were engineered in to the
HBc DNA sequence after amino acid residue position
149. After digestion with EcoRI and HindIII, a
synthetic DNA having the above AATT 5'overhang
followed by a T cell epitope-encoding sequence, a
stop codon and a 3' AGCT overhang were ligated into
the digested sequence to form a sequence that encoded
HBc residues 1-149 followed by two heterologous
residues (GI), the stop codon and the HindIII site.

PCR amplification using a forward primer
having a SacI restriction site followed by a sequence
encoding HBc beginning at residue position 79,

followed by digestion with SacI and HindIII provided
a sequence encoding HBc positions 79-149 plus the two
added residues and the T cell epitope at the C-
terminus. Digestion of the construct of Fig. 2B with
SacI and ligation provided the complete gene encoding
a desired recombinant HBc chimer immunogen having the
sequence, from the N-terminus, of HBc positions 1-78,
two added residues, the malarial B cell epitope, two
added residues, HBc positions 79-149, two added
residues, and the T cell epitope that is shown in
Fig. 2C.

It is noted that the preferred use of two
heterologous residues on either side of (flanking) a
B cell or T cell epitope is a matter of convenience.
As a consequence, one can also use zero to three or
more added residues that are not part of the HBc
sequence on either or both sides of an inserted
sequence. One or both ends of the insert and HBc
nucleic acid can be "chewed back" with an appropriate
nuclease (e.g. S1 nuclease) to provide blunt ends
that can be ligated together. Added heterologous
residues that are neither part of the inserted B cell
or T cell epitopes nor a part of the HBc sequence are
not counted in the number of residues present in a
recited Domain.

It is also noted that one can also
synthesize all or a part of a desired recombinant HBc
chimer nucleic acid using well-known synthetic
methods as is discussed and illustrated in U. S.
Patent No.5,656,472 for the synthesis of the 177 base
pair DNA that encodes the 59 residue ribulose bis-
phosphate carboxylase-oxygenase signal peptide of
Nicotiana tabacum. For example, one can synthesize
Domains I and II with a blunt or "sticky" end that

can be ligated to Domains III and IV to provide a
construct that expresses a contemplated HBc chimer
that contains zero added residues to the N-terminal
side of the B cell epitope and zero to three added
residues on the C-terminal side or at the Domain
II/III junction or at some other desired location.

A nucleic acid sequence (segment) that
encodes a previously described HBc chimer molecule or
a complement of that coding sequence is also
contemplated herein. Such a nucleic acid segment is
present in isolated and purified form in some
preferred embodiments.

In living organisms, the amino acid residue
sequence of a protein or polypeptide is directly
related via the genetic code to the deoxyribonucleic
acid (DNA) sequence of the gene that codes for the
protein. Thus, through the well-known degeneracy of
the genetic code additional DNAs and corresponding
RNA sequences (nucleic acids) can be prepared as
desired that encode the same chimer amino acid
residue sequences, but are sufficiently different
from a before-discussed gene sequence that the two
sequences do not hybridize at high stringency, but do
hybridize at moderate stringency.

High stringency conditions can be defined
as comprising hybridization at a temperature of about
50°-55°C in 6XSSC and a final wash at a temperature of
68°C in 1-3XSSC. Moderate stringency conditions
comprise hybridization at a temperature of about 50°C
to about 65°C in 0.2 to 0.3 M NaCl, followed by
washing at about 50°C to about 55°C in 0.2X SSC, 0.1%
SDS (sodium dodecyl sulfate).

A nucleic sequence (DNA sequence or an RNA
sequence) that (1) itself encodes, or its complement

encodes, a chimer molecule whose HBc portion from
residue position 1 through 136, when present, is that
of SEQ ID NOs: 168, 170, 171, 172, 173 or 174 and (2)
hybridizes with a DNA sequence of SEQ ID NOs: 169,
175, 176, 177, 178 or 179 at least at moderate
stringency (discussed above); and (3) whose HBc
sequence shares at least 80 percent, and more
preferably at least 90 percent, and even more
preferably at least 95 percent, and most preferably
100 percent identity with a DNA sequence of SEQ ID
NOs: 169, 175, 176, 177, 178 and 179, is defined as a
DNA variant sequence. As is well-known, a nucleic
acid sequence such as a contemplated nucleic acid
sequence is expressed when operatively linked to an
appropriate promoter in an appropriate expression
system as discussed elsewhere herein.

        An analog or analogous nucleic acid (DNA or
RNA) sequence that encodes a contemplated chimer
molecule is also contemplated as part of this
invention. A chimer analog nucleic acid sequence or
its complementary nucleic acid sequence encodes a HBc
amino acid residue sequence that is at least 80
percent, and more preferably at least 90 percent, and
most preferably is at least 95 percent identical to
the HBc sequence portion from residue position 1
through residue position 136 shown in SEQ ID NOs:
168, 170, 171, 172, 173 and 174. This DNA or RNA is
referred to herein as an "analog of" or "analogous
to" a sequence of a nucleic acid of SEQ ID NOs: 169,
175, 176, 177, 178 and 179, and hybridizes with the
nucleic acid sequence of SEQ ID NOs: 169, 175, 176,
177, 178 and 179 or their complements herein under
moderate stringency hybridization conditions. A
nucleic acid that encodes an analogous sequence, upon

suitable transfection and expression, also produces a
contemplated chimer.

        Different hosts often have preferences for
a particular codon to be used for encoding a
particular amino acid residue. Such codon
preferences are well known and a DNA sequence
encoding a desired chimer sequence can be altered,
using *in vitro* mutagenesis for example, so that host-
preferred codons are utilized for a particular host
in which the enzyme is to be expressed. In addition,
one can also use the degeneracy of the genetic code
to encode the HBc portion of a sequence of SEQ ID
NOs: 168, 170, 171, 172, 173 or 174 that avoids
substantial identity with a DNA of SEQ ID Nos: 169,
175, 176, 177, 178 or 179, or their complements.
Thus, a useful analogous DNA sequence need not
hybridize with the nucleotide sequences of SEQ ID
NOs: 169, 175, 176, 177, 178 or 179 or a complement
under conditions of moderate stringency, but can
still provide a contemplated chimer molecule.

        A recombinant nucleic acid molecule such as
a DNA molecule, comprising a vector operatively
linked to an exogenous nucleic acid segment (e.g., a
DNA segment or sequence) that defines a gene that
encodes a contemplated chimer, as discussed above,
and a promoter suitable for driving the expression of
the gene in a compatible host organism, is also
contemplated in this invention. More particularly,
also contemplated is a recombinant DNA molecule that
comprises a vector comprising a promoter for driving
the expression of the chimer in host organism cells
operatively linked to a DNA segment that defines a
gene for the HBc portion of a chimer or a DNA variant
that has at least 90 percent identity to the chimer

gene of SEQ ID NOs: 169, 175, 176, 177, 178 or 179
and hybridizes with that gene under moderate
stringency conditions.

Further contemplated is a recombinant DNA
molecule that comprises a vector containing a
promoter for driving the expression of a chimer in
host organism cells operatively linked to a DNA
segment that is an analog nucleic acid sequence that
encodes an amino acid residue sequence of a HBc
chimer portion that is at least 80 percent identical,
more preferably 90 percent identical, and most
preferably 95 percent identical to the HBc portion of
a sequence of SEQ ID NOs: 168, 170, 171, 172, 173 or
174. That recombinant DNA molecule, upon suitable
transfection and expression in a host cell, provides
a contemplated chimer molecule.

It is noted that because of the 30 amino
acid residue N-terminal sequence of ground squirrel
HBc does not align with any of the other HBc
sequences, that sequence and its encoding nucleic
acid sequences and their complements are not included
in the above percentages of identity, nor are the
portions of nucleic acid that encode that 30-residue
sequence or its complement used in hybridization
determinations. Similarly, sequences that are
truncated at either or both of the HBc N- and C-
termini are not included in identity calculations,
nor are those sequences in which residues of the
immunodominant loop are removed for insertion of a
heterologous epitope. Thus, only those HBc-encoding
bases or HBc sequence residues that are present in a
chimer molecule are included and compared to an
aligned nucleic acid or amino acid residue sequence
in the identity percentage calculations.

Inasmuch as the coding sequences for the
gene disclosed herein is illustrated in SEQ ID NOs:
169, 175, 176, 177, 178 and 179, isolated nucleic
acid segments, preferably DNA sequences, variants and
analogs thereof can be prepared by *in vitro*
mutagenesis, as is well known in the art and
discussed in Current Protocols In Molecular Biology,
Ausabel et al. eds., John Wiley & Sons (New York:
1987) p. 8.1.1-8.1.6, that begin at the initial ATG
codon for a gene and end at or just downstream of the
stop codon for each gene. Thus, a desired
restriction site can be engineered at or upstream of
the initiation codon, and at or downstream of the
stop codon so that other genes can be prepared,
excised and isolated.

As is well known in the art, so long as the
required nucleic acid, illustratively DNA sequence,
is present, (including start and stop signals),
additional base pairs can usually be present at
either end of the segment and that segment can still
be utilized to express the protein. This, of course,
presumes the absence in the segment of an operatively
linked DNA sequence that represses expression,
expresses a further product that consumes the enzyme
desired to be expressed, expresses a product that
consumes a wanted reaction product produced by that
desired enzyme, or otherwise interferes with
expression of the gene of the DNA segment.

Thus, so long as the DNA segment is free of
such interfering DNA sequences, a DNA segment of the
invention can be about 500 to about 15,000 base pairs
in length. The maximum size of a recombinant DNA
molecule, particularly an expression vector, is
governed mostly by convenience and the vector size

that can be accommodated by a host cell, once all of the minimal DNA sequences required for replication and expression, when desired, are present. Minimal vector sizes are well known. Such long DNA segments are not preferred, but can be used.

DNA segments that encode the before-described chimer can be synthesized by chemical techniques, for example, the phosphotriester method of Matteucci et al. (1981) *J. Am. Chem. Soc.*, **103**:3185. Of course, by chemically synthesizing the coding sequence, any desired modifications can be made simply by substituting the appropriate bases for those encoding the native amino acid residue sequence. However, DNA segments including sequences discussed previously are preferred.

A contemplated HBc chimer can be produced (expressed) in a number of transformed host systems, typically host cells although expression in acellular, *in vitro*, systems is also contemplated. These host cellular systems include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g. baculovirus); plant cell systems transformed with virus expression vectors (e.g. cauliflower mosaic virus; tobacco mosaic virus) or with bacterial expression vectors (e.g., Ti plasmid); or appropriately transformed animal cell systems such as CHO or COS cells. The invention is not limited by the host cell employed.

DNA segments containing a gene encoding the HBc chimer are preferably obtained from recombinant

DNA molecules (plasmid vectors) containing that gene. Vectors capable of directing the expression of a chimer gene into the protein of a HBc chimer is referred to herein as an "expression vector".

An expression vector contains expression control elements including the promoter. The chimer-coding gene is operatively linked to the expression vector to permit the promoter sequence to direct RNA polymerase binding and expression of the chimer-encoding gene. Useful in expressing the polypeptide coding gene are promoters that are inducible, viral, synthetic, constitutive as described by Poszkowski et al. (1989) *EMBO J.*, 3:2719 and Odell et al. (1985) *Nature*, 313:810, as well as temporally regulated, spatially regulated, and spatiotemporally regulated as given in Chua et al. (1989) *Science*, 244:174-181.

One preferred promoter for use in prokaryotic cells such as *E. coli* is the Rec 7 promoter that is inducible by exogenously supplied nalidixic acid. A more preferred promoter is present in plasmid vector JHEX25 (available from Promega) that is inducible by exogenously supplied isopropyl-β-D-thiogalacto-pyranoside (IPTG). A still more preferred promoter, the tac promoter, is present in plasmid vector pKK223-3 and is also inducible by exogenously supplied IPTG. The pKK223-3 plasmid can be successfully expressed in a number of *E. coli* strains, such as XL-1, TB1, BL21 and BLR, using about 25 to about 100 µM IPTG for induction. Surprisingly, concentrations of about 25 to about 50 µM IPTG have been found to provide optimal results in 2 L shaker flasks and fermentors.

Several strains of *Salmonella* such as *S. typhi* and *S. typhimurium* and *S. typhimurium-E. coli*

hybrids have been used to express immunogenic transgenes including prior HBc chimer particles both as sources of the particles for use as immunogens and as live, attenuated whole cell vaccines and inocula, and those expression and vaccination systems can be used herein. See, U.S. Patent No. 6,024,961; U.S. Patent No. 5,888,799; U.S. Patent No. 5,387,744; U.S. Patent No. 5,297,441; Ulrich et al., (1998) *Adv. Virus Res.*, 50:141-182; Tacket et al., (Aug 1997) *Infect. Immun.*, 65(8):3381-3385; Schodel et al., (Feb 1997) *Behring Inst. Mitt.*, 98:114-119; Nardelli-Haefliger et al., (Dec 1996) *Infect. Immun.*, 64(12):5219-5224; Londono et al., (Apr 1996) *Vaccine,* 14(6):545-552, and the citations therein.

Expression vectors compatible with eukaryotic cells, such as those compatible with yeast cells or those compatible with cells of higher plants or mammals, are also contemplated herein. Such expression vectors can also be used to form the recombinant DNA molecules of the present invention. Vectors for use in yeasts such as *S. cerivisiae* or *Pichia pastoris* can be episomal or integrating, as is well known. Eukaryotic cell expression vectors are well known in the art and are available from several commercial sources. Normally, such vectors contain one or more convenient restriction sites for insertion of the desired DNA segment and promoter sequences. Optionally, such vectors contain a selectable marker specific for use in eukaryotic cells. Exemplary promoters for use in *S. cerevisiae* include the *S. cerevisiae* phosphoglyceric acid kinase (PGK) promoter and the divergent promoters GAL 10 and GAL 1, whereas the alcohol oxidase gene (AOX1) is a useful promoter for *Pichia pastoris*.

For example, to produce chimers in the methylotrophic yeast, *P. pastoris*, a gene that encodes a desired chimer is placed under the control of regulatory sequences that direct expression of structural genes in *Pichia*. The resultant expression-competent forms of those genes are introduced into *Pichia* cells.

More specifically, the transformation and expression system described by Cregg et al. (1987) *Biotechnology*, 5:479-485; (1987) *Molecular and Cellular Biology*, 12:3376-3385 can be used. A gene for a chimer V12.Pf3.1 is placed downstream from the alcohol oxidase gene (AOX1) promoter and upstream from the transcription terminator sequence of the same AOX1 gene. The gene and its flanking regulatory regions are then introduced into a plasmid that carries both the *P. pastoris* HIS4 gene and a *P. pastoris* ARS sequence (Autonomously Replicating Sequence), which permit plasmid replication within *P. pastoris* cells [Cregg et al. (1987) *Molecular and Cellular Biology*, 12:3376-3385].

The vector also contains appropriate portions of a plasmid such as pBR322 to permit growth of the plasmid in *E. coli* cells. The resultant plasmid carrying a chimer gene, as well as the various additional elements described above, is illustratively transformed into a his4 mutant of *P. pastoris*; i.e. cells of a strain lacking a functional histidinol dehydrogenase gene.

After selecting transformant colonies on media lacking histidine, cells are grown on media lacking histidine, but containing methanol as described Cregg et al. (1987) *Molecular and Cellular Biology*, 12:3376-3385, to induce the AOX1 promoters.

The induced AOX1 promoters cause expression of the chimer protein and the production of chimer particles in *P. pastoris*.

A contemplated chimer gene can also be introduced by integrative transformation, which does not require the use of an ARS sequence, as described by Cregg et al. (1987) *Molecular and Cellular Biology*, **12**:3376-3385.

Production of chimer particles by recombinant DNA expression in mammalian cells is illustratively carried out using a recombinant DNA vector capable of expressing the chimer gene in Chinese hamster ovary (CHO) cells. This is accomplished using procedures that are well known in the art and are described in more detail in Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratories (1989).

In one illustrative example, the simian virus (SV40) based expression vector, pKSV-10 (Pharmacia Fine Chemicals, Piscataway, NJ), is subjected to restriction endonuclease digestion by NcoI and HindIII. A NcoI/HindIII sequence fragment that encodes the desired HBc chimer prepared as described in Example 1 is ligated into the expression plasmid, which results in the formation of a circular recombinant expression plasmid denominated pSV-Pf.

The expression plasmid pSV-Pf contains an intact *E. coli* ampicillin resistance gene. *E. coli* RR101 (Bethesda Research Laboratories, Gaithersburg, MD), when transformed with pSV-Pf, can thus be selected on the basis of ampicillin resistance for those bacteria containing the plasmid. Plasmid-containing bacteria are then cloned and the clones are subsequently screened for the proper orientation

of the inserted coding gene into the expression   ·
vector.

 The above obtained plasmid, pSV-Pf,
containing the gene that encodes a desired HBc chimer
is propagated by culturing *E. coli* containing the
plasmid. The plasmid DNA is isolated from *E. coli*
cultures as described in Sambrook et al., above.

 Expression of a chimer is accomplished by
the introduction of pSV-Pf into the mammalian cell
line, e.g., CHO cells, using the calcium phosphate-
mediated transfection method of Graham et al.(1973)
Virol., **52**:456, or a similar technique.

 To help ensure maximal efficiency in the
introduction of pSV-Pf into CHO cells in culture, the
transfection is carried out in the presence of a
second plasmid, pSV2NEO (ATCC #37149) and the
cytotoxic drug G418 (GIBCO Laboratories, Grand
Island, N.Y.) as described by Southern et al. (1982)
J. Mol. Appl. Genet., **1**:327. Those CHO cells that
are resistant to G418 are cultured, have acquired
both plasmids, pSV2NEO and pSV-Pf, and are designated
CHO/pSV-Pf cells. By virtue of the genetic
architecture of the pSV-Pf expression vector, a
chimer is expressed in the resulting CHO/pSV-Pf cells
and can be detected in and purified from the
cytoplasm of these cells. The resulting composition
containing cellular protein is separated on a column
as discussed elsewhere herein.

 The choice of which expression vector and
ultimately to which promoter a chimer-encoding gene
is operatively linked depends directly on the
functional properties desired, e.g. the location and
timing of protein expression, and the host cell to be
transformed. These are well known limitations

inherent in the art of constructing recombinant DNA molecules. However, a vector useful in practicing the present invention can direct the replication, and preferably also the expression (for an expression vector) of the chimer gene included in the DNA segment to which it is operatively linked.

In one preferred embodiment, the host that expresses the chimer is the prokaryote, *E. coli*, and a preferred vector includes a prokaryotic replicon; i.e., a DNA sequence having the ability to direct autonomous replication and maintenance of the recombinant DNA molecule extrachromosomally in a prokaryotic host cell transformed therewith. Such replicons are well known in the art.

Those vectors that include a prokaryotic replicon can also include a prokaryotic promoter region capable of directing the expression of a contemplated HBc chimer gene in a host cell, such as *E. coli*, transformed therewith. Promoter sequences compatible with bacterial hosts are typically provided in plasmid vectors containing one or more convenient restriction sites for insertion of a contemplated DNA segment. Typical of such vector plasmids are pUC8, pUC9, and pBR329 available from Biorad Laboratories, (Richmond, CA) and pPL and pKK223-3 available from Pharmacia, Piscataway, NJ.

Typical vectors useful for expression of genes in cells from higher plants and mammals are well known in the art and include plant vectors derived from the tumor-inducing (Ti) plasmid of *Agrobacterium tumefaciens* described by Rogers et al. (1987) *Meth. in Enzymol.*, **153**:253-277 and mammalian expression vectors pKSV-10, above, and pCI-neo (Promega Corp., #E1841, Madison, WI). However,

several other expression vector systems are known to
function in plants including pCaMVCN transfer control
vector described by Fromm et al. (1985) *Proc. Natl.
Acad. Sci. USA*, 82:58-24. Plasmid pCaMVCN (available
from Pharmacia, Piscataway, NJ) includes the
cauliflower mosaic virus CaMV 35S promoter.

The above plant expression systems
typically provide systemic or constitutive expression
of an inserted transgene. Systemic expression can be
useful where most or all of a plant is used as the
source to a contemplated chimer molecule or resultant
particles or where a large part of the plant is used
to provide an oral vaccine. However, it can be more
efficacious to express a chimer molecule or particles
in a plant storage organ such as a root, seed or
fruit from which the particles can be more readily
isolated or ingested.

One manner of achieving storage organ
expression is to use a promoter that expresses its
controlled gene in one or more preselected or
predetermined non-photosynthetic plant organs.
Expression in one or more preselected storage organs
with little or no expression in other organs such as
roots, seed or fruit versus leaves or stems is
referred to herein as enhanced or preferential
expression. An exemplary promoter that directs
expression in one or more preselected organs as
compared to another organ at a ratio of at least 5:1
is defined herein as an organ-enhanced promoter.
Expression in substantially only one storage organ
and substantially no expression in other storage
organs is referred to as organ-specific expression;
i.e., a ratio of expression products in a storage
organ relative to another of about 100:1 or greater

indicates organ specificity. Storage organ-specific
promoters are thus members of the class of storage
organ-enhanced promoters.

Exemplary plant storage organs include the
roots of carrots, taro or manioc, potato tubers, and
the meat of fruit such as red guava, passion fruit,
mango, papaya, tomato, avocado, cherry, tangerine,
mandarin, palm, melons such cantaloupe and
watermelons and other fleshy fruits such as squash,
cucumbers, mangos, apricots, peaches, as well as the
seeds of maize (corn), soybeans, rice, oil seed rape
and the like.

The CaMV 35S promoter is normally deemed to
be a constitutive promoter. However, recent research
has shown that a 21-bp region of the CaMV 35S
promoter, when operatively linked into another,
heterologous usual green tissue promoter, the rbcS-3A
promoter, can cause the resulting chimeric promoter
to become a root-enhanced promoter. That 21-bp
sequence is disclosed in U.S. Patent No. 5,023,179.
The chimeric rbcS-3A promoter containing the 21-bp
insert of U.S. Patent No. 5,023,179 is a useful root-
enhanced promoter herein.

A similar root-enhanced promoter, that
includes the above 21-bp segment is the -90 to +8
region of the CAMV 35S promoter itself. U.S. Patent
No. 5,110,732 discloses that that truncated CaMV 35S
promoter provides enhanced expression in roots and
the radical of seed, a tissue destined to become a
root. That promoter is also useful herein.

Another useful root-enhanced promoter is
the -1616 to -1 promoter of the oil seed rape
(*Brassica napus L.*) gene disclosed in PCT/GB92/00416
(WO 91/13922 published Sep. 19, 1991). *E. coli*

DH5.alpha. harboring plasmid pRlambdaS4 and bacteriophage lambda.beta.1 that contain this promoter were deposited at the National Collection of Industrial and Marine Bacteria, Aberdeen, GB on Mar. 8, 1990 and have accession numbers NCIMB40265 and NCIMB40266. A useful portion of this promoter can be obtained as a 1.0 kb fragment by cleavage of the plasmid with HaeIII.

A preferred root-enhanced promoter is the mannopine synthase (mas) promoter present in plasmid pKan2 described by DiRita and Gelvin (1987) *Mol. Gen. Genet*, 207:233-241. This promoter is removable from its plasmid pKan2 as a XbaI-XbalI fragment.

The preferred mannopine synthase root-enhanced promoter is comprised of the core mannopine synthase (mas) promoter region up to position -138 and the mannopine synthase activator from -318 to -213, and is collectively referred to as AmasPmas. This promoter has been found to increase production in tobacco roots about 10- to about 100-fold compared to leaf expression levels.

Another root specific promoter is the about 500 bp 5' flanking sequence accompanying the hydroxyproline-rich glycopeprotein gene, HRGPnt3, expressed during lateral root initiation and reported by Keller et al. (1989) *Genes Dev.*, 3:1639-1646. Another preferred root-specific promoter is present in the about -636 to -1 5' flanking region of the tobacco root-specific gene ToRBF reported by Yamamoto et al. (1991) *Plant Cell*, 3:371-381. The cis-acting elements regulating expression are more specifically located by those authors in the region from about -636 to about -299 5' from the transcription initiation site. Yamamoto et al. reported steady

state mRNA production from the ToRBF gene in roots, but not in leaves, shoot meristems or stems.

Still another useful storage organ-specific promoter are the 5' and 3' flanking regions of the fruit-ripening gene E8 of the tomato, *Lycopersicon esculentum*. These regions and their cDNA sequences are illustrated and discussed in Deikman et al. (1988) *EMBO J.*, 7(11):3315-3320 and (1992) *Plant Physiol.*, 100:2013-2017.

Three regions are located in the 2181 bp of the 5' flanking sequence of the gene and a 522 bp sequence 3' to the poly (A) addition site appeared to control expression of the E8 gene. One region from -2181 to -1088 is required for activation of E8 gene transcription in unripe fruit by ethylene and also contributes to transcription during ripening. Two further regions, -1088 to -863 and -409 to -263, are unable to confer ethylene responsiveness in unripe fruit but are sufficient for E8 gene expression during ripening.

The maize sucrose synthase-1 (Sh) promoter that in corn expresses its controlled enzyme at high levels in endosperm, at much reduced levels in roots and not in green tissues or pollen has been reported to express a chimeric reporter gene, β-glucuronidase (GUS), specifically in tobacco phloem cells that are abundant in stems and roots. Yang et al. (1990) *Proc. Natl. Acad. Sci., U.S.A.*, 87:4144-4148. This promoter is thus useful for plant organs such as fleshy fruits like melons, e.g. cantaloupe, or seeds that contain endosperm and for roots that have high levels of phloem cells.

Another exemplary tissue-specific promoter is the lectin promoter, which is specific for seed

tissue. The lectin protein in soybean seeds is encoded by a single gene (Le1) that is only expressed during seed maturation and accounts for about 2 to about 5 percent of total seed mRNA. The lectin gene and seed-specific promoter have been fully characterized and used to direct seed specific expression in transgenic tobacco plants. See, e.g., Vodkin et al. (1983) *Cell*, 34:1023 and Lindstrom et al. (1990) *Developmental Genetics*, 11:160.

A particularly preferred tuber-specific expression promoter is the 5' flanking region of the potato patatin gene. Use of this promoter is described in Twell et al. (1987) *Plant Mol. Biol.*, 9:365-375. This promoter is present in an about 406 bp fragment of bacteriophage LPOTI. The LPOTI promoter has regions of over 90 percent homology with four other patatin promoters and about 95 percent homology over all 400 bases with patatin promoter PGT5. Each of these promoters is useful herein. See, also, Wenzler et al. (1989) *Plant Mol. Biol.*, 12:41-50.

Still further organ-enhanced and organ-specific promoter are disclosed in Benfey et al. (1988) *Science*, 244:174-181.

Each of the promoter sequences utilized is substantially unaffected by the amount of chimer molecule or particles in the cell. As used herein, the term "substantially unaffected" means that the promoter is not responsive to direct feedback control (inhibition) by the chimer molecules or particles accumulated in transformed cells or transgenic plant.

Transfection of plant cells using *Agrobacterium tumefaciens* is typically best carried out on dicotyledonous plants. Monocots are usually

most readily transformed by so-called direct gene
transfer of protoplasts. Direct gene transfer is
usually carried out by electroportation, by
polyethyleneglycol-mediated transfer or bombardment
of cells by microprojectiles carrying the needed DNA.
These methods of transfection are well-known in the
art and need not be further discussed herein.
Methods of regenerating whole plants from transfected
cells and protoplasts are also well-known, as are
techniques for obtaining a desired protein from plant
tissues. See, also, U.S. Patents No. 5,618,988 and
5,679,880 and the citations therein.

A transgenic plant formed using
Agrobacterium transformation, electroportation or
other methods typically contains a single gene on one
chromosome. Such transgenic plants can be referred to
as being heterozygous for the added gene. However,
inasmuch as use of the word "heterozygous" usually
implies the presence of a complementary gene at the
same locus of the second chromosome of a pair of
chromosomes, and there is no such gene in a plant
containing one added gene as here, it is believed
that a more accurate name for such a plant is an
independent segregant, because the added, exogenous
chimer molecule-encoding gene segregates
independently during mitosis and meiosis. A
transgenic plant containing an organ-enhanced
promoter driving a single structural gene that
encodes a contemplated HBc chimeric molecule; i.e.,
an independent segregant, is a preferred transgenic
plant.

More preferred is a transgenic plant that
is homozygous for the added structural gene; i.e., a
transgenic plant that contains two added genes, one

gene at the same locus on each chromosome of a
chromosome pair. A homozygous transgenic plant can
be obtained by sexually mating (selfing) an
independent segregant transgenic plant that contains
a single added gene, germinating some of the seed
produced and analyzing the resulting plants produced
for enhanced chimer particle accumulation relative to
a control (native, non-transgenic) or an independent
segregant transgenic plant. A homozygous transgenic
plant exhibits enhanced chimer particle accumulation
as compared to both a native, non-transgenic plant
and an independent segregant transgenic plant.

It is to be understood that two different
transgenic plants can also be mated to produce
offspring that contain two independently segregating
added, exogenous (heterologous) genes. Selfing of
appropriate progeny can produce plants that are
homozygous for both added, exogenous genes that
encode a chimeric HBc molecule. Back-crossing to a
parental plant and out-crossing with a non-transgenic
plant are also contemplated.

A transgenic plant of this invention thus
has a heterologous structural gene that encodes a
contemplated chimeric HBc molecule. A preferred
transgenic plant is an independent segregant for the
added heterologous chimeric HBc structural gene and
can transmit that gene to its progeny. A more
preferred transgenic plant is homozygous for the
heterologous gene, and transmits that gene to all of
its offspring on sexual mating.

Inasmuch as a gene that encodes a chimeric
HBc molecule does not occur naturally in plants, a
contemplated transgenic plant accumulates chimeric
HBc molecule particles in a greater amount than does

-64-

a non-transformed plant of the same type or strain when both plants are grown under the same conditions.

The phrase "same type" or "same strain" is used herein to mean a plant of the same cross as or a clone of the untransformed plant. Where alleic variations among siblings of a cross are small, as with extensively inbred plant, comparisons between siblings can be used or an average arrived at using several siblings. Otherwise, clones are preferred for the comparison.

Seed from a transgenic plant is grown in the field greenhouse, window sill or the like, and resulting sexually mature transgenic plants are self-pollinated to generate true breeding plants. The progeny from these plants become true breeding lines that are evaluated for chimeric HBc molecule particle accumulation, preferably in the field, under a range of environmental conditions.

A transgenic plant homozygous for chimeric HBc molecule particle accumulation is crossed with a parent plant having other desired traits. The progeny, which are heterozygous or independently segregatable for chimeric HBc molecule particle accumulation, are backcrossed with one or the other parent to obtain transgenic plants that exhibit chimeric HBc molecule particle accumulation and the other desired traits. The backcrossing of progeny with the parent may have to be repeated more than once to obtain a transgenic plant that possesses a number of desirable traits.

An insect cell system can also be used to express a HBc chimer. For example, in one such system *Autographa californica* nuclear polyhedrosis virus (AcNPV) or baculovirus is used as a vector to

express foreign genes in *Spodoptera frugiperda* cells or in Trichoplusia larvae.

The sequences encoding a chimer can be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of chimer sequence renders the polyhedrin gene inactive and produces recombinant virus lacking coat protein. The recombinant viruses can then be used to infect, for example, *S. Frugiperda* cells or *Trichoplusia* larvae in which the HBc chimer can be expressed. E. Engelhard et al. (1994) *Proc. Natl. Acad. Sci., USA,* 91:3224-3227; and V. Luckow, Insect Cell Expression Technology, pp. 183-218, in Protein Engineering: Principles and Practice, J.L. Cleland et al. eds., Wiley-Liss, Inc, 1996). Heterologous genes placed under the control of the polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcNPV) are often expressed at high levels during the late stages of infection.

Recombinant baculoviruses containing the chimeric gene are constructed using the baculovirus shuttle vector system (Luckow et al. (1993) *J. Virol.,* 67:4566-4579], sold commercially as the Bac-To-Bac™ baculovirus expression system (Life Technologies). Stocks of recombinant viruses are prepared and expression of the recombinant protein is monitored by standard protocols (O'Reilly et al., Baculovirus Expression Vectors: A Laboratory Manual, W.H. Freeman and Company, New York, 1992; and King et al., The Baculovirus Expression System: A Laboratory Guide, Chapman & Hall, London, 1992).

A variety of methods have been developed to operatively link DNA to vectors via complementary

cohesive termini or blunt ends. For instance, complementary homopolymer tracts can be added to the DNA segment to be inserted into the vector DNA. The vector and DNA segment are then joined by hydrogen bonding between the complementary homopolymeric tails to form recombinant DNA molecules.

Alternatively, synthetic linkers containing one or more restriction endonuclease sites can be used to join the DNA segment to the expression vector, as noted before. The synthetic linkers are attached to blunt-ended DNA segments by incubating the blunt-ended DNA segments with a large excess of synthetic linker molecules in the presence of an enzyme that is able to catalyze the ligation of blunt-ended DNA molecules, such as bacteriophage T4 DNA ligase.

Thus, the products of the reaction are DNA segments carrying synthetic linker sequences at their ends. These DNA segments are then cleaved with the appropriate restriction endonuclease and ligated into an expression vector that has been cleaved with an enzyme that produces termini compatible with those of the synthetic linker. Synthetic linkers containing a variety of restriction endonuclease sites are commercially available from a number of sources including New England BioLabs, Beverly, MA. A desired DNA segment can also be obtained using PCR technology in which the forward and reverse primers contain desired restriction sites that can be cut after amplification so that the gene can be inserted into the vector. Alternatively PCR products can be directly cloned into vectors containing T-overhangs (Promega Corp., A3600, Madison, WI) as is well known in the art.

The expressed chimeric protein self-
assembles into particles within the host cells,
whether in single cells or in cells within a
multicelled host. The particle-containing cells are
harvested using standard procedures, and the cells
are lysed using a French pressure cell, lysozyme,
sonicator, bead beater or a microfluidizer
(Microfluidics International Corp., Newton MA).
After clarification of the lysate, particles are
precipitated with 45% ammonium sulfate, resuspended
in 20 mM sodium phosphate, pH 6.8 and dialyzed
against the same buffer. The dialyzed material is
clarified by brief centrifugation and the supernatant
subjected to gel filtration chromatography using
Sepharose® CL-4B. Particle-containing fractions are
identified, subjected to hydroxyapatite
chromatography, and reprecipitated with ammonium
sulfate prior to resuspension, dialysis and sterile
filtration and storage at -70°C.

Malarial Inocula and Vaccines

A before-described recombinant HBc chimer
immunogen preferably in particulate form is dissolved or
dispersed in an immunogenic effective amount in a
pharmaceutically acceptable vehicle composition that is
preferably aqueous to form an inoculum or vaccine. When
administered to a host animal in need of immunization or
in which antibodies are desired to be induced such as a
mammal (e.g., a mouse, dog, goat, sheep, horse, bovine,
monkey, ape, or human) or bird (e.g., a chicken, turkey,
duck or goose), an inoculum induces antibodies that
immunoreact with the malarial B cell epitope present in
the immunogen. In a vaccine, those induced antibodies
also immunoreact in vivo with (bind to) the sporozoite and

protect the mammal from malarial infection by the
Plasmodium species whose B cell epitope was present in the
immunogen. A composition that is an inoculum in one
animal can be a vaccine for another where the *Plasmodium*
species against which antibodies are raised does not
infect the animal inoculated, as where an inoculum against
P. falciparum is used to raise antibodies in mice.

The amount of recombinant HBc chimer
immunogen utilized in each immunization is referred
to as an immunogenic effective amount and can vary
widely, depending *inter alia*, upon the recombinant
HBc chimer immunogen, mammal immunized, and the
presence of an adjuvant in the vaccine, as discussed
below. Immunogenic effective amounts for a vaccine
and an inoculum provide the protection or antibody
activity, respectively, discussed hereinbefore.

Vaccines or inocula typically contain a
recombinant HBc chimer immunogen concentration of about 1
microgram to about 1 milligram per inoculation (unit
dose), and preferably about 10 micrograms to about 50
micrograms per unit dose. Immunizations in mice typically
contain 10 or 20 µg of chimer particles.

The term "unit dose" as it pertains to a vaccine
or inoculum of the present invention refers to a
physically discrete unit suitable as an unitary dosage for
animals, each unit containing a predetermined quantity of
active material calculated to individually or collectively
produce the desired immunogenic effect in association with
the required diluent; i.e., carrier, or vehicle. A single
unit dose or a plurality of unit doses can be used to
provide an immunogenic effective amount of recombinant HBc
chimer immunogen.

Vaccines or inocula are typically prepared
from a recovered recombinant HBc chimer immunogen by

dispersing the immunogen in a physiologically
tolerable (acceptable) diluent vehicle such as water,
saline phosphate-buffered saline (PBS), acetate-
buffered saline (ABS), Ringer's solution or the like
to form an aqueous composition. The diluent vehicle
can also include oleaginous materials such as peanut
oil, squalane or squalene as is discussed
hereinafter.

The immunogenic active ingredient is often
mixed with excipients that are pharmaceutically
acceptable and compatible with the active ingredient.
Suitable excipients are, for example, water, saline,
dextrose, glycerol, ethanol, or the like and
combinations thereof. In addition, if desired, an
inoculum or vaccine can contain minor amounts of
auxiliary substances such as wetting or emulsifying
agents, pH buffering agents that enhance the
immunogenic effectiveness of the composition.

A contemplated vaccine or inoculum
advantageously also includes an adjuvant. Suitable
adjuvants for vaccines and inocula of the present
invention comprise those adjuvants that are capable
of enhancing the antibody responses against B cell
epitopes of the chimer, as well as adjuvants capable
of enhancing cell mediated responses towards T cell
epitopes contained in the chimer. Adjuvants are well
known in the art (see, for example, Vaccine Design -
The Subunit and Adjuvant Approach, 1995,
Pharmaceutical Biotechnology, Volume 6, Eds. Powell,
M.F., and Newman, M.J., Plenum Press, New York and
London, ISBN 0-306-44867-X).

Exemplary adjuvants include complete
Freund's adjuvant (CFA) that is not used in humans,
incomplete Freund's adjuvant (IFA), squalene,

squalane and alum [e.g., Alhydrogel™ (Superfos, Denmark)], which are materials well known in the art, and are available commercially from several sources.

Preferred adjuvants for use with immunogens of the present invention include aluminum or calcium salts (for example hydroxide or phosphate salts). A particularly preferred adjuvant for use herein is an aluminum hydroxide gel such as Alhydrogel™. For aluminum hydroxide gels (alum), the chimer protein is admixed with the adjuvant so that between 50 to 800 micrograms of aluminum are present per dose, and preferably between 400 and 600 micrograms are present.

Another particularly preferred adjuvant for use with an immunogen of the present invention is an emulsion. A contemplated emulsion can be an oil-in-water emulsion or a water-in-oil emulsions. In addition to the immunogenic chimer protein, such emulsions comprise an oil phase of squalene, squalane, peanut oil or the like as are well-known, and a dispersing agent. Non-ionic dispersing agents are preferred and such materials include mono- and di-C_{12}-C_{24}-fatty acid esters of sorbitan and mannide such as sorbitan mono-stearate, sorbitan mono-oleate and mannide mono-oleate. An immunogen-containing emulsion is administered as an emulsion.

Preferably, such emulsions are water-in-oil emulsions that comprise squalene and mannide mono-oleate (Arlacel™ A), optionally with squalane, emulsified with the chimer protein in an aqueous phase. Well-known examples of such emulsions include Montanide™ ISA-720, and Montanide™ ISA 703 (Seppic, Castres, France), each of which is understood to contain both squalene and squalane, with squalene

predominating in each, but to a lesser extent in
Montanide™ ISA 703. Most preferably, Montanide™ ISA-
720 is used, and a ratio of oil-to-water of 7:3 (w/w)
is used. Other preferred oil-in-water emulsion
adjuvants include those disclosed in WO 95/17210 and
EP 0 399 843.

The use of small molecule adjuvants is also
contemplated herein. One type of small molecule
adjuvant useful herein is a 7-substituted-8-oxo- or
8-sulfo-guanosine derivative described in U.S.
Patents No. 4,539,205, No. 4,643,992, No. 5,011,828
and No. 5,093,318, whose disclosures are incorporated
by reference. Of these materials, 7-allyl-8-
oxoguanosine (loxoribine) is particularly preferred.
That molecule has been shown to be particularly
effective in inducing an antigen-(immunogen-)specific
response.

Still further useful adjuvants include
monophosphoryl lipid A (MPL) available from Corixa
Corp. (see, U.S. Patent No. 4,987,237), CPG available
from Coley Pharmaceutical Group, QS21 available from
Aquila Biopharmaceuticals, Inc., SBAS2 available from
SKB, the so-called muramyl dipeptide analogues
described in U.S. Patent No. 4,767,842, and MF59
available from Chiron Corp. (see, U.S. Patents No.
5,709,879 and No. 6,086,901).

More particularly, immunologically active
saponin fractions having adjuvant activity derived
from the bark of the South American tree *Quillaja
Saponaria Molina* (e.g. Quil™ A) are also useful.
Derivatives of Quil™ A, for example QS21 (an HPLC
purified fraction derivative of Quil™ A), and the
method of its production is disclosed in U.S. Patent

No.5,057,540. In addition to QS21 (known as QA21), other fractions such as QA17 are also disclosed.

3-De-O-acylated monophosphoryl lipid A is a well-known adjuvant manufactured by Ribi Immunochem, Hamilton, Montana. The adjuvant contains three components extracted from bacteria, monophosphoryl lipid (MPL) A, trehalose dimycolate (TDM) and cell wall skeleton (CWS) (MPL+TDM+CWS) in a 2% squalene/Tween® 80 emulsion. This adjuvant can be prepared by the methods taught in GB 2122204B. A preferred form of 3-de-O-acylated monophosphoryl lipid A is in the form of an emulsion having a small particle size less than 0.2 μm in diameter (EP 0 689 454 B1).

The muramyl dipeptide adjuvants include N-acetyl-muramyl-L-threonyl-D-isoglutamine(thur-MDP), N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP), and N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmityol-sn-glycero-3-hydroxyphosphoryloxy)-ethylamin (CGP) 1983A, referred to as MTP-PE).

Preferred adjuvant mixtures include combinations of 3D-MPL and QS21 (EP 0 671 948 B1), oil-in-water emulsions comprising 3D-MPL and QS21 (WO 95/17210, PCT/EP98/05714)., 3D-MPL formulated with other carriers (EP 0 689 454 B1), QS21 formulated in cholesterol-containing liposomes (WO 96/33739), or immunostimulatory oligonucleotides (WO 96/02555). Alternative adjuvants include those described in WO 99/52549 and non-particulate suspensions of polyoxyethylene ether (UK Patent Application No. 9807805.8).

Adjuvants are utilized in an adjuvant amount, which can vary with the adjuvant, mammal and

recombinant HBc chimer immunogen. Typical amounts
can vary from about 1 µg to about 1 mg per
immunization. Those skilled in the art know that
appropriate concentrations or amounts can be readily
determined.

Inocula and vaccines are conventionally
administered parenterally, by injection, for example,
either subcutaneously or intramuscularly. Additional
formulations that are suitable for other modes of
administration include suppositories and, in some
cases, oral formulation or by nasal spray. For
suppositories, traditional binders and carriers can
include, for example, polyalkalene glycols or
triglycerides; such suppositories may be formed from
mixtures containing the active ingredient in the
range of 0.5% to 10%, preferably 1-2%. Oral
formulations include such normally employed
excipients as, for example, pharmaceutical grades of
mannitol, lactose, starch, magnesium stearate, sodium
saccharine, cellulose, magnesium carbonate and the
like.

An inoculum or vaccine composition takes
the form of a solution, suspension, tablet, pill,
capsule, sustained release formulation or powder, and
contains an immunogenic effective amount of HBc
chimer, preferably as particles, as active
ingredient. In a typical composition, an immunogenic
effective amount of preferred HBc chimer particles is
about 1 µg to about 1 mg of active ingredient per
dose, and more preferably about 5 µg to about 50 µg
per dose, as noted before.

A vaccine or inoculum is typically
formulated for parenteral administration. Exemplary
immunizations are carried out sub-cutaneously (SC)

intra-muscularly (IM), intravenusly (IV), intraperitoneally (IP) or intra-dermally (ID).

The HBc chimer particles and HBc chimer particle conjugates can be formulated into the vaccine as neutral or salt forms. Pharmaceutically acceptable salts, include the acid addition salts (formed with the free amino groups of the protein or hapten) and are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived form inorganic bases such as, for example, sodium, potassium, ammonium,· calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

The inocula or vaccines are administered in a manner compatible with the dosage formulation, and in such amount as are therapeutically effective and immunogenic. The quantity to be administered depends on the subject to be treated, capacity of the subject's immune system to synthesize antibodies, and degree of protection desired. Precise amounts of active ingredient required to be administered depend on the judgment of the practitioner and are peculiar to each individual. However, suitable dosage ranges are of the order of several hundred micrograms active ingredient per individual. Suitable regimes for initial administration and booster shots are also variable, but are typified by an initial administration followed in intervals (weeks or months) by a subsequent injection or other administration.

Once immunized, the mammal is maintained
for a period of time sufficient for the recombinant
HBc chimer immunogen to induce the production of a
sufficient titer of antibodies that bind to
sporozoite. The maintenance time for the production
of anti-sporozoite antibodies typically lasts for a
period of about three to about twelve weeks, and can
include a booster, second immunizing administration
of the vaccine. A third immunization is also
contemplated, if desired, at a time 24 weeks to five
years after the first immunization. It is
particularly contemplated that once a protective
level titer of anti-sporozoite antibodies is
attained, that the vaccinated mammal is preferably
maintained at or near that antibody titer by periodic
booster immunizations administered at intervals of
about 1 to about 5 years.

The production of anti-sporozoite
antibodies is readily ascertained by obtaining a
plasma or serum sample from the immunized mammal and
assaying the antibodies therein for their ability to
bind to a synthetic circumsporozoite immunodominant
antigen [e.g. the *P. falciparum* CS protein peptide
(NANP)$_5$ used herein] in an ELISA assay as described
hereinafter or by another immunoassay such as a
Western blot as is well known in the art. Most
preferable is the use of the indirect
immunofluorescence assay (IFA), in which intact
sporozoites are employed as the capture antigen,
discussed hereinafter.

It is noted that the before-described anti-
CS antibodies so induced can be isolated from the
blood of the host mammal using well known techniques,
and then reconstituted into a second vaccine for

passive immunization as is also well known. Similar
techniques are used for gamma-globulin immunizations
of humans. For example, antiserum from one or a
number of immunized hosts can be precipitated in
aqueous ammonium sulfate (typically at 40-50 percent
of saturation), and the precipitated antibodies
purified chromatographically as by use of affinity
chromatography in which (NANP)$_5$ is utilized as the
antigen immobilized on the chromatographic column.

Inocula are preparations that are
substantially identical to vaccines, but are used in
a host mammal in which antibodies to malaria are
desired to be induced, but in which protection from
malaria is not desired. In one example, a vaccine
against *P. falciparum* of *P. vivax* can be used in mice
as an inoculum to induce antibody production and not
be a vaccine because neither malarial species can
infect mice. Similarly, a similar inoculum can be
used in a horse or sheep to induce antibody
production against either or both malarial species
for use in a passive immunization in yet another
animal such as humans.

Best Mode for Carrying Out the Invention

Without further elaboration, it is believed
that one skilled in the art can, using the preceding
description and the detailed examples below, utilize
the present invention to its fullest extent. The
following preferred specific embodiments are,
therefore, to be construed as merely illustrative,
and not limiting of the remainder of the disclosure
in any way whatsoever.

Example 1: B Cell Epitope-Containing

-77-

Chimer Preparation _____

A. Preparation of plasmid vector pKK223-
    3N, a modified form of pKK223-3 _____

Plasmid vector pKK223-3 (Pharmacia) was
modified by the establishment of a unique NcoI
restriction site to enable insertion of HBc genes as
NcoI-HindIII restriction fragments and subsequent
expression in *E.coli* host cells. To modify the
pKK223-3 plasmid vector, a new SphI-HindIII fragment
was prepared using the PCR primers pKK223-3/433-452-F
and pKK223-NcoI-mod-R, and pKK223-3 as the template.

This PCR fragment was cut with the
restriction enzymes SphI and HindIII to provide a 467
bp fragment that was then ligated with a 4106 bp
fragment of the pKK223-3 vector, to effectively
replace the original 480 bp SphI-HindIII fragment.
The resultant plasmid (pKK223-3N) is therefore 13 bp
shorter than the parent plasmid and contains modified
nucleotide sequence upstream of the introduced NcoI
site (see Fig. 1 in which the dashes indicate the
absent bases). The final plasmid, pKK223-3N, has a
size of 4573 bp. Restriction sites in plasmid
pKK223-3N are indicated in Fig. 1, and the nucleotide
changes made to pKK223-3 to form plasmid pKK223-3N
are indicated by an underline as shown below.

pKK223-3/433-452-F   GGTGCATGCAAGGAGATG   SEQ ID NO:27

pKK223-NcoI-mod-R
GCGAAGCTTCGGATCccatggTTTTTTCCTCCTTATGTGAAATTGTTATCCG-
              CTC   SEQ ID NO:28

B. Preparation of V1
    and V2 Cloning Vectors

Modified HBc149 genes, able to accept the
directional insertion of synthetic dsDNA fragments
into the immunodominant loop region, were constructed
using PCR. (The plasmid accepting inserts between
amino acids E77 and D78 was named V1, whereas the
plasmid accepting inserts between D78 and P79 was
named V2.) The HBc149 gene was amplified in two
halves using two PCR primer pairs, one of which
amplifies the amino terminus, the other amplifies the
carboxyl terminus. For V1, the products of the PCR
reactions (N- and C-terminus) are both 246 bp
fragments; for V2, the products are a 249 bp (N-
terminus) and a 243 bp fragment (C-terminus).

The N-terminal fragments prepared were
digested with NcoI and EcoRI, and the C-terminal
fragments were digested with EcoRI and HindIII. The
V1 and V2 fragments pairs were then ligated together
at the common EcoRI overhangs. The resultant NcoI-
HindIII fragments were then ligated into the pKK223-
3N vector, which had been prepared by digestion with
NcoI and HindIII.

To insert B cell epitopes into the V1 and
V2 plasmids, the plasmids were digested with EcoRI
and SacI restriction enzymes. Synthetic dsDNA
fragments containing 5' EcoRI and 3' SacI overhangs
were then inserted. In both cases, V1 and V2,
glycine-isoleucine (EcoRI) and glutamic acid-leucine
(SacI) amino acid pairs, coded for by the restriction
sites, flank the inserted B cell epitopes. The
inserted restriction sites are underlined in the
primers below.

V1

HBc149/NcoI-F

5'-TTGGG<u>CCATGG</u>ACATCGACCCTTA   SEQ ID NO:29

HBc-E77/EcoRI-R

5'-GCG<u>GAATTC</u>CTTCCAAATTAACACCCACC   SEQ ID NO:30

HBc-D78/EcoRI-SacI-F

5'-CGC<u>GAATTC</u>AAAAA<u>GAGCTC</u>GATCCAGCGTCTAGAGAC

                                            SEQ ID NO:31

HBc149/HindIII-R

5'-CGC<u>AAGCTT</u>AAACAACAGTAGTCTCCGGAAG   SEQ ID NO:32

<u>V2</u>

HBc149/NcoI-F

5'-TTGGG<u>CCATGG</u>ACATCGACCCTTA   SEQ ID NO:29

HBc-D78/EcoRI-R

5'-GCG<u>GAATTC</u>CATCTTCCAAATTAACACCCAC   SEQ ID NO:186

HBc-P79/EcoRI-SacI-F

5'-CGC<u>GAATTC</u>AAAAA<u>GAGCTC</u>CCAGCGTCTAGAGACCTAG

                                            SEQ ID NO:34

HBc149/HindIII-R

5'-CGC<u>AAGCTT</u>AAACAACAGTAGTCTCCGGAAG   SEQ ID NO:32

C. <u>Preparation of V7 Cloning Vector</u>

To enable the fusion of T cell epitopes to
the C terminus of a HBc chimer, a new vector, V7, was
constructed. Unique EcoRI and SacI restriction sites

were inserted between valine-149 and the HindIII site
to facilitate directional insertion of synthetic
dsDNAs into EcoRI-HindIII (or EcoRI-SacI) restriction
sites. The pair of PCR primers below was used to
amplify the HBc 149 gene with a NcoI restriction site
at the amino-terminus and EcoRI, SacI and HindIII
sites at the carboxyl-terminus. The product of the
PCR reaction (479 bp) was digested with NcoI/HindIII
and cloned into pKK223-3N to form V7.

     To insert T cell epitopes, the plasmid (V7)
was digested EcoRI/HindIII (or EcoRI-SacI) and
synthetic dsDNA fragments having EcoRI/HindIII (or
EcoRI/SacI) overhangs, were ligated into V7. For all
V7 constructs, the final amino acid of native HBc
(valine-149) and the first amino acid of the inserted
T cell epitope are separated by a glycine-isoleucine
dipeptide sequence coded for by the nucleotides that
form the EcoRI restriction site. For epitopes
inserted at EcoRI/SacI, there are additional glutamic
acid-leucine residues after the T cell epitope, prior
to the termination codon, contributed by the SacI
site. Restriction sites are again underlined in the
primers shown.

HBc149/NcoI-F

5'-TTGGGCCATGGACATCGACCCTTA   SEQ ID NO:29

HBc149/SacI-EcoRI-H3-R

5'-CGCAAGCTTAGAGCTCTTGAATTCCAACAACAGTAGTCTCCG

                                          SEQ ID NO:33

     D.   Preparation of V12
          Expression Constructs

V12 vectors, which contain B cell epitopes between amino acids 78 and 79, as well as T cell epitopes downstream of valine-149, were constructed from V2 and V7 vectors. The carboxyl terminus of a V7 vector containing a T cell epitope inserted at EcoRI/HindIII was amplified using two PCR primers (HBc-P79/SacI-F and pKK223-2/4515-32R) to provide a dsDNA fragment corresponding to amino acids 79-149 plus the T cell epitope, flanked with SacI and HindIII restriction sites.

The PCR products were cut with SacI and HindIII and then cloned into the desired V2 vector prepared by cutting with the same two enzymes. The PCR primers shown are amenable for the amplification of the carboxyl terminus of all V7 genes, irrespective of the T cell epitope present after amino acid 149 of the HBc gene.

One exception to the generality of this approach was in the preparation of the V12 constructs containing the Pf-CS(C17A) mutation, which were prepared from existing V12 constructs. In this case, V12 constructs were amplified with HBc149/NcoI-F (SEQ ID NO:29) and the mis-match reverse PCR primer (SEQ ID NO: 104), which facilitated the C17A mutation. The resultant PCR product was digested with NcoI and HindIII and cloned back into pKK223-3N (previously cut with the same enzymes). Restriction sites are underlined.

HBc-P79/SacI-F   5'-CGCGAGCTCCCAGCGTCTAGAGACCTAG

                                                SEQ ID NO:35

pKK223-2/4515-32R   5'-GTATCAGGCTGAAAATC

E. *P.falciparum* CS-repeat B cell
Epitopes Inserted into V2

For V2 and V7 constructs, synthetic dsDNA
fragments coding for the B (V2) or T cell epitope
(V7) of interest were inserted into EcoRI/SacI
restriction sites. Synthetic dsDNA fragments,
encoding B and T cell epitopes of interest, were
prepared by mixing complementary single stranded DNA
oligonucleotides at equimolar concentrations, heating
to 95°C for 5 minutes, and then cooling to room
temperature at a rate of -1 °C per minute. This
annealing reaction was performed in TE buffer. The
double-stranded DNAs are shown below with the encoded
epitope sequence shown above. The pound symbol, #,
is used in some of the amino acid residue sequences
that follow to indicate the presence of a stop codon.

Pf1

```
    I   N   A   N   P   N   A   N   P   N   A   N   P   N   A
AATTAACGCTAATCCGAACGCTAATCCGAACGCTAATCCGAACGCTA
    TTGCGATTAGGCTTGCGATTAGGCTTGCGATTAGGCTTGCGAT
```

N P E L	SEQ ID NO: 37
ATCCGGAGCT	SEQ ID NO: 38
TAGGCC	SEQ ID NO: 39

Pf3

   I N A N P N V D P N A N P N A N P

AATTAACGCTAATCCGAACGTTGACCCGAACGCTAATCCGAACGCTAATCCGA

  TTGCGATTAGGCTTGCAACTGGGCTTGCGATTAGGCTTGCGATTAGGCT

N A N P N V D P N A N P E L SEQ ID NO:40

ACGCTAATCCGAACGTTGACCCGAACGCTAATCCGGAGCT SEQ ID NO:41

TGCGATTAGGCTTGCAACTGGGCTTGCGATTAGGCCTCGAGG

                                             SEQ ID NO:42

Pf3.1

   I N A N P N V D P N A N P N A N P

AATTAACGCGAATCCGAACGTGGATCCGAATGCCAACCCTAACGCCAACCC

  TTGCGCTTAGGCTTGCACCTAGGCTTACGGTTGGGATTGCGGTTGGG

             N A N P E L   SEQ ID NO:43

           AAATGCGAACCCAGAGCT   SEQ ID NO:44

           TTTACGCTTGGGTC   SEQ ID NO:45

Pf3.2

   I N A N P N A N P N A N P N V D P

AATTAACGCGAATCCGAATGCCAACCCTAACGCCAACCCAAACGTGGATCCGA

  TTGCGCTTAGGCTTACGGTTGGGATTGCGGTTGGGTTTGCACCTAGGCT

            N A N P E L   SEQ ID NO:46

           ATGCGAACCCAGAGCT   SEQ ID NO:47

           TACGCTTGGGTC   SEQ ID NO:48

Pf3.3


```
  I   N   A   N   P   N   V   D   P   N   A   N   P   N   A   N   P
AATTAACGCGAATCCGAACGTGGATCCAAATGCCAACCCTAACGCTAATCCAA
   TTGCGCTTAGGCTTGCACCTAGGTTTACGGTTGGGATTGCGATTAGGTT
```

```
  N   A   N   P   N   V   D   PN   A   N   P   E   L    SEQ ID NO:49
ACGCCAACCCGAATGTTGACCCCAATGCCAATCCGGAGCT  SEQ ID NO:50
TGCGGTTGGGCTTACAACTGGGGTTACGGTTAGGCC      SEQ ID NO:51
```


Pf3.4


```
  I   N   P   N   V   D   P   N   A   N   P   N   A   N   P   N   A
AATTAATCCGAACGTGGATCCAAATGCCAACCCTAACGCTAATCCAAACGCCA
   TTAGGCTTGCACCTAGGTTTACGGTTGGGATTGCGATTAGGTTTGCGGT
```

```
        N   P   N   V   E   L                SEQ ID NO:52
        ACCCGAATGTTGAGCT                     SEQ ID NO:53
        TGGGCTTACAAC                         SEQ ID NO:54
```


Pf3.5


```
  I   N   P   N   V   D   P   N   A   N   P   N   A   N   P   N   A
AATTAATCCGAACGTGGATCCAAATGCCAACCCTAACGCTAATCCAAACGCCA
   TTAGGCTTGCACCTAGGTTTACGGTTGGGATTGCGATTAGGTTTGCGGT
```

```
        N   P   N   V   D   P   E   L        SEQ ID NO:55
        ACCCGAATGTTGACCCTGAGCT               SEQ ID NO:56
        TGGGCTTACAACTGGGAC                   SEQ ID NO:57
```


Pf3.6

```
     I  N  P  N  V  D  P  N  A  N  P  N  A  N  P  N  A
  AATTAATCCGAACGTGGATCCAAATGCCAACCCTAACGCTAATCCAAACGCCA
  TTAGGCTTGCACCTAGGTTTACGGTTGGGATTGCGATTAGGTTTGCGGT
```

```
            N  P  N  V  D  P  N  A  E  L    SEQ ID NO:58
         ACCCGAATGTTGACCCTAATGCTGAGCT        SEQ ID NO:59
         TGGGCTTACAACTGGGATTACGAC            SEQ ID NO:60
```

Pf3.7

```
     I  N  V  D  P  N  A  N  P  N  A  N  P  N  A  N  P
  AATTAACGTGGATCCAAATGCCAACCCTAACGCTAATCCAAACGCCAACCCGA
  TTGCACCTAGGTTTACGGTTGGGATTGCGATTAGGTTTGCGGTTGGGCT
```

```
            N  V  E  L           SEQ ID NO:61
         ATGTTGAGCT              SEQ ID NO:62
         TACAAC                  SEQ ID NO:63
```

Pf3.8

```
     I  N  V  D  P  N  A  N  P  N  A  N  P  N  A  N  P
  AATTAACGTGGATCCAAATGCCAACCCTAACGCTAATCCAAACGCCAACCCGA
  TTGCACCTAGGTTTACGGTTGGGATTGCGATTAGGTTTGCGGTTGGGCT
```

```
            N  V  D  P  E  L       SEQ ID NO:64
         ATGTTGACCCTGAGCT          SEQ ID NO:65
         TACAACTGGGAC              SEQ ID NO:66
```

Pf3.9

  I N V D P N A N P N A N P N A N P

AATTAACGTGGATCCAAATGCCAACCCTAACGCTAATCCAAACGCCAACCCGA

TTGCACCTAGGTTTACGGTTGGGATTGCGATTAGGTTTGCGGTTGGGCT

     N V D P N A E L   SEQ ID NO:67

    ATGTTGACCCTAATGCTGAGCT   SEQ ID NO:68

    TACAACTGGGATTACGAC   SEQ ID NO:69

Pf3.10

  I D P N A N P N A N P N A N P

AATTGATCCAAATGCCAACCCTAACGCTAATCCAAACGCCAACC

CTAGGTTTACGGTTGGGATTGCGATTAGGTTTGCGGTTGG

      N V E L   SEQ ID NO:70

    CGAATGTTGAGCT   SEQ ID NO:71

    GCTTACAAC   SEQ ID NO:72

Pf3.11

  I D P N A N P N A N P N A N P N V

AATTGATCCAAATGCCAACCCTAACGCTAATCCAAACGCCAACCCGAATGTTG

CTAGGTTTACGGTTGGGATTGCGATTAGGTTTGCGGTTGGGCTTACAAC

     D P E L   SEQ ID NO:73

    ACCCTGAGCT   SEQ ID NO:74

    TGGGAC   SEQ ID NO:75

Pf3.12

    I   D   P   N   A   N   P   N   A   N   P   N   A   N   P   N   V
AATTGATCCAAATGCCAACCCTAACGCTAATCCAAACGCCAACCCGAATGTTG
    CTAGGTTTACGGTTGGGATTGCGATTAGGTTTGCGGTTGGGCTTACAAC

        D   P   N   A   E   L   SEQ ID NO:76
        ACCCTAATGCCGAGCT   SEQ ID NO:77
        TGGGATTACGGC   SEQ ID NO:78

        F.   *P.falciparum* universal T cell epitope

Pf-UTC (PF/CS326-345)

    I   E   Y   L   N   K   I   Q   N   S   L   S   T   E   W   S   P
AATTGAATATCTGAACAAAATCCAGAACTCTCTGTCCACCGAATGGTCTCCGT
    CTTATAGACTTGTTTTAGGTCTTGAGAGACAGGTGGCTTACCAGAGGCA

        C   S   V   T   #   #   SEQ ID NO:79
        GCTCCGTTACCTAGTA   SEQ ID NO:80
        CGAGGCAATGGATCATTCGA   SEQ ID NO:81
        P.vivax CS-repeat B cell epitopes

Pv-T1A

    I   P   A   G   D   R   A   D   G   Q   P   A   G   D   R   A   A
AATTCCGGCTGGTGACCGTGCAGATGGCCAGCCAGCGGGTGACCGCGCTGCAG
    GGCCGACCACTGGCACGTCTACCGGTCGGTCGCCCACTGGCGCGACGTC

        G   Q   P   A   G   E   L   SEQ ID NO:82
        GCCAGCCGGCTGGCGAGCT   SEQ ID NO:83
        CGGTCGGCCGACCGC   SEQ ID NO:84

Pv-T1B

```
    I   D   R   A   A   G   Q   P   A   G   D   R   A   D   G   Q   P
AATTGACAGAGCAGCCGGACAACCAGCAGGCGATCGAGCAGACGGACAGCCCG
    CTGTCTCGTCGGCCTGTTGGTCGTCCGCTAGCTCGTCTGCCTGTCGGGC
```

```
            A   G   E   L                       SEQ ID NO:85
            CAGGGGAGCT                           SEQ ID NO:86
            GTCCCC                               SEQ ID NO:87
```

Pv-T2A

```
    I   A   N   G   A   G   N   Q   P   G   A   N   G   A   G   D   Q
AATTGCGAACGGCGCCGGTAATCAGCCGGGGGGCAAACGGCGCGGGTGATCAAC
    CGCTTGCCGCGGCCATTAGTCGGCCCCCGTTTGCCGCGCCCACTAGTTG
```

```
            P   G   E   L                       SEQ ID NO:88
            CAGGGGAGCT                           SEQ ID NO:89
            GTCCCC                               SEQ ID NO:90
```

Pv-T2B

```
    I   A   N   G   A   D   N   Q   P   G   A   N   G   A   D   D   Q
AATTGCGAACGGCGCCGATAATCAGCCGGGTGCAAACGGGGCGGATGACCAAC
    CGCTTGCCGCGGCTATTAGTCGGCCCACGTTTGCCCCGCCTACTGGTTG
```

```
            P   G   E   L                       SEQ ID NO:91
            CAGGCGAGCT                           SEQ ID NO:92
            GTCCGC                               SEQ ID NO:93
```

Pv-T2C

```
      I  A  N  G  A  G  N  Q  P  G  A  N  G  A  G  D  Q
  AATTGCGAACGGCGCCGGTAATCAGCCGGGAGCAAACGGCGCGGGGGATCAAC
      CGCTTGCCGCGGCCATTAGTCGGCCCTCGTTTGCCGCGCCCCCTAGTTG


   P  G  A  N  G  A  D  N  Q  P  G  A  N  G  A  D  D
  CAGGCGCCAATGGTGCAGACAACCAGCCTGGGGCGAATGGAGCCGATGACC
  GTCCGCGGTTACCACGTCTGTTGGTCGGACCCCGCTTACCTCGGCTACTGG
```

```
              Q   P   G   E   L              SEQ ID NO:94
              AACCCGGCGAGCT                  SEQ ID NO:95
              TTGGGCCGC                      SEQ ID NO:96
```

PV-T3

```
      I  A  P  G  A  N  Q  E  G  G  A  A  A  P  G  A  N
  AATTGCGCCGGGCGCCAACCAGGAAGGTGGGGCTGCAGCGCCAGGAGCCAATC
      CGCGGCCCGCGGTTGGTCCTTCCACCCCGACGTCGCGGTCCTCGGTTAG
```

```
              Q   E   G   G   A   A   E   L      SEQ ID NO:97
              AAGAAGGCGGTGCAGCGGAGCT              SEQ ID NO:98
              TTCTTCCGCCACGTCGCC                  SEQ ID NO:99
```

Example 2: _P.vivax universal T cell epitope_

Pv-UTC

```
      I  E  Y  L  D  K  V  R  A  T  V  G  T  E  W  T  P
  AATTGAATATCTGGATAAAGTGCGTGCGACCGTTGGCACGGAATGGACTCCGT
      CTTATAGACCTATTTCACGCACGCTGGCAACCGTGCCTTACCTGAGGCA
```

```
              C   S   V   T   #   #          SEQ ID NO:100
              GCAGCGTGACCTAATA               SEQ ID NO:101
              CGTCGCACTGGATTATTCGA           SEQ ID NO:102
```

A. PCR primers for site-directed
    mutagenesis

Pf-CS(C17A)-R   SEQ ID NO:103

     # # T V S A P S W E T S
GCC<u>AAGCTT</u>ACTAGGTAACGGAGGCCGGAGACCATTCGGTGG
     HindIII   SEQ ID NO:104

                    B. PCR Primers for Truncation and
                        Cysteine Addition at C-terminus
         To modify the C-terminus of HBc chimer
genes, either via the addition of cysteine residues
or varying the length of the HBc gene, PCR reactions
were performed using HBc149 as template with the
HBc/NcoI-F primer and a reverse primer (e.g.
HBc149+C/HindIII-R) that directed the desired
modification of the C-terminus. PCR products were
digested with NcoI and HindIII (whose restriction
sites are underlined), and cloned into pKK223-3N at
the same restriction sites.

HBc149/NcoI-F   SEQ ID NO:105
            M   D   I   D   P   Y
5'-TTGGG<u>CCATGG</u>ACATCGACCCTTA   SEQ ID NO:29

HBc149+C/HindIII-R   SEQ ID NO:106
          # # C V V T T E P L
5'-CGC<u>AAGCTT</u>ACTAGCAAACAACAGTAGTCTCCGGAAG

                                           SEQ ID NO:107

HBc144/HindIII-R   SEQ ID NO:108
    # P   L   T   S   L   I   P
CGC<u>AAGCTT</u>ACGGAAGTGTTGATAGGATAGGG   SEQ ID NO:109

HBc142/HindIII-R   SEQ ID NO:110
     # T   S   L   I   P   A   N   P
CGC<u>AAGCTT</u>ATGTTGATAGGATAGGGGCATTTGG   SEQ ID NO:111

HBc140/HindIII-R   SEQ ID NO:112
    # L   I   P   A   N   P   P
CGC<u>AAGCTT</u>ATAGGATAGGGGCATTTGGTGG   SEQ ID NO:113

HBc139/HindIII-R   SEQ ID NO:114
    # I   P   A   N   P   P
GCG<u>AAGCTT</u>AGATAGGGGCATTTGGTGG   SEQ ID NO:115

HBc138/HindIII-R   SEQ ID NO:116
    # P   A   N   P   P   R
CGC<u>AAGCTT</u>AAGGGGCATTTGGTGGTCT   SEQ ID NO:117

HBc138+C/HindIII-R   SEQ ID NO:118
    # C   P   A   N   P   P   R
GCG<u>AAGCTT</u>AGCAAGGGGCATTTGGTGGTCT   SEQ ID NO:119

HBc137/HindIII-R   SEQ ID NO:120
    # A   N   P   P   R   Y   A
GCG<u>AAGCTT</u>AGGCATTTGGTGGTCTATAGC   SEQ ID NO:121

HBc137+C/HindIII-R   SEQ ID NO:122
   # C A   N   P   P   R   Y   A
GCG<u>AAGCTT</u>AGCAGGCATTTGGTGGTCTATAA   SEQ ID NO:123

HBc136/HindIII-R   SEQ ID NO:124

     # N P P R Y A P

CGCAAGCTTAATTTGGTGGTCTATAAGCTGG   SEQ ID NO:125

Example 3: Assay Procedures

    A. Antigenicity

    1. Particle ELISA

Purified particles were diluted to a concentration of 10 µg/mL in coating buffer (50 mM sodium bicarbonate, pH 9.6) and coated onto the wells of ELISA strips (50 µL/well). The ELISA strips were incubated at room temperature overnight (about 18 hours). Next morning, the wells were washed with ELISA wash buffer [phosphate buffered saline (PBS), pH 7.4, 0.05% Tween®-20] and blocked with 3% BSA in PBS for 1 hour (75 µL/well). ELISA strips were stored, dry, at -20°C until needed.

To determine the antigenicity of particles, antisera were diluted using 1% BSA in PBS and 50 µL/well added to antigen-coated ELISA wells. Sera were incubated for 1 hour, washed with ELISA wash buffer (above) and probed using an anti-mouse (IgG)-HRP (The Binding Site, San Diego, CA; HRP = horseradish peroxidase) conjugate (50 µL/well) or other appropriate antibody for 30 minutes. After washing with ELISA wash buffer the reaction was visualized by the addition of TM blue substrate (50 µL/well). After 10 minutes, the reaction was stopped by the addition of 1N H_2SO_4 (100 µL/well) and read on an ELISA plate reader set at 450 nm.

2. Synthetic Peptide ELISA

A 20 amino acid residue synthetic peptide (NANP)$_5$ was diluted to a concentration of 2 µg/mL in coating buffer (50 mM sodium bicarbonate, pH 9.6) and coated onto the wells of ELISA strips (50 µL/well). Peptides were dried onto the wells by incubating overnight (about 18 hours), in a hood with the exhaust on. Next morning, the wells were washed with ELISA wash buffer (phosphate buffered saline, pH 7.4, 0.05% Tween®-20) and blocked with 3% BSA in PBS (75 µL/well) for 1 hour. ELISA strips were stored, dry, at -20°C until needed.

To determine antibody antigenicity of particles, antisera (monoclonal or polyclonal) were diluted using 1% BSA in PBS, and 50 µL/well added to antigen-coated ELISA wells. Sera were incubated for 1 hour, washed with ELISA wash buffer, and probed using an anti-mouse(IgG)-HRP conjugate or other antibody (as above at 50 µL/well) for 30 minutes, washed again with ELISA wash buffer, and then visualized by the addition of TM blue substrate (50 µL/well). After 10 minutes, the reaction was stopped by the addition of 1N H_2SO_4 (100 µL/well) and read on an ELISA plate reader set at 450 nm.

B. Immunogenicity of Particles

To assay the immunogenicity of particles, mice were immunized, IP, with 20 µg of particles in Freund's complete adjuvant, and then boosted at 4 weeks with 10 µg in Freund's incomplete adjuvant. Mice were bled at 2, 4, 6, and 8 weeks.

C. Sporozoite IFA

Indirect immunofluorescence assay (IFA) was carried out using glutaraldehyde-fixed *P. falciparum* sporozoites and FITC-labeled anti-mouse IgG (gamma-chain specific) (Kirkegaard and Perry, Gaithersburg, MD) to detect bound antibody [Munesinghe *et al.*, *Eur.J.Immunol.* 1991, 21, 3015-3020]. Sporozoites used were dissected from the salivary glands of *Anopheles* mosquitoes infected by feeding on *P.falciparum* (NF54 isolate) gametocytes derived from *in vitro* cultures.

Example 4: Expression of Recombinant
            Chimer HBc Particles

A. Effect of Insertion
    Position on Immunogenicity

Antibody titers (1/reciprocal dilution) were measured for mice immunized with HBc particles containing the *P. f*-CS B cell epitope (NANP)$_4$, inserted either between amino acids E77/D78 or D78/P79, or by using a loop replacement approach (CS-2) [discussed in Schodel et al., (1994) *J. Exp. Med.*, 180:1037-1046, using complete Freund's adjuvant]. Mice were immunized with a single 20 μg dose, IP, with adjuvant as noted before, and antibody titers determined in an ELISA using immobilized (NANP)$_5$ synthetic peptide. The results of those studies are shown in Table 1, below.

Table 1

Time	CS-2*	E77/D78 (V1)	D78/P79 (V2)
2 weeks	0	2,560	2,560
4 weeks	640	2,560	40,960

*Schodel et al., (1994) *J.Exp.Med.*, 180:1037-1046.

Another comparison was made of insertion position of the NANP CS-repeat epitope on immunogenicity, using BALB/c mice. Antibody titers induced by the CS-2 particle of Schodel et al. were compared to titers achieved using the same $(NANP)_4$ B cell epitope, inserted between HBc positions 78 and 79, and using the above V2.Pf1 particles as immunogen. Sera were analyzed 4 weeks after primary (1°) and 2 weeks after booster (2°) immunization, and the results are shown in Table 2, below.

Table 2

Chimer	Primary	Booster
CS-2	0	640*
V2.Pf1	10,240	655,360

* Schodel et al., (1994) *J.Exp.Med.*, 180:1037-1046

A similar comparison of insertion position of the NANP CS-repeat epitope on immunogenicity was made using B10.S mice, and the results are shown in Table 3.

Table 3

Chimer	Primary	Booster
CS-2	640*	20,480*
V2.Pf1	163,840	655,360

* Schodel et al., (1994) *J.Exp.Med.*, 180:1037-1046

The effect on the immunogenicity of HBc chimer particles (ELISA, F1 mice) that include the minor B cell epitope, NANPNVDP (SEQ ID NO:126) that includes the sequence NVDP (SEQ ID NO: 185), along with a repeated NANP sequence was examined. A HBc chimer was expressed that contained the sequence NANPNVDP(NANP)$_3$NVDP (SEQ ID NO:2; V12.Pf3) inserted between HBc positions 78 and 79. The resulting ELISA data were compared to titers obtained using the tetrameric repeat (NANP)$_4$ B cell epitope (V12.Pf1) or the dimer of the minor B cell epitope at the same position (V12.Pf7). Each of these three chimers contained a Domain IV that included the HBc sequence from position 141 through 149, bonded to the *P. falciparum* universal T cell epitope as the C-terminal sequence. The results of these studies using primary and booster immunizations as discussed before and using adjuvants, are shown below in Table 4.

Table 4

Chimer	Primary	Booster
V12.Pf1	163,840	655,360
V12.Pf3	2,621,440	10,485,760
V12.Pf7	2,560	

The observed greater than 20-fold increase in immunogenicity by including the 'minor' repeat epitope was quite unexpected. Because V12.Pf3 was

not well expressed by *E.coli*, variants of the Pf3 epitope NANPNVDP(NANP)$_3$NVDP (SEQ ID NO:2) were constructed that had similar antigenicity to Pf3, but with increased expression levels, as shown below. Only constructs 3.1 and 3.2 were assayed for immunogenicity.

Relative expression levels of recombinant chimer HBc/*P. falciparum* particles and antigenicities for monoclonal antibodies specific for the CS epitopes (NANP)$_4$ and (NANPNVDP) are shown in Table 5 below. Relative expression levels are as follows; ****=75-125 mg/L; ***=50-75 mg/L; **=25-50 mg/L. Antigenicity was determined by end point titer dilutions for the monoclonal antibodies [MoAb 2A10 for (NANP)$_4$; MoAb 2B6D8 for NANPNVDP; and *P. vivax* Rpt. MoAb 2F2 were provided by E. Nardin of New York University Medical Center]. The data were normalized such that the lowest titer is expressed as 1. For example, V12.Pf3 was 165 fold more antigenic than V12.Pf3.10 for the (NANP)$_4$-specific monoclonal, and 26-fold more antigenic than V12.Pf3.2 for the NANPNVDP-specific monoclonal antibody. N.D.= no detectable antibody binding. [Note: V12.Pf3.7 was not expressed due to a mutation in the expression vector; it was not examined further because similar constructs were not antigenic, and re-cloning was therefore not a worthwhile endeavor.]

Table 5

Name	*P.falciparum* B Cell Epitope	Relative Expression	Antigenicity (NANP)$_4$	NANPNVDP
V12.Pf1	(NANP)$_4$ SEQ ID NO:1	****	33	ND
V12.Pf3	NANPNVDP(NANP)$_3$NVDP SEQ ID NO:2	**	165	31
V12.Pf3.1	NANPNVDP(NANP)$_3$ SEQ ID NO:3	****	33	31
V12.Pf3.2	(NANP)$_3$NVDPNANP SEQ ID NO:4	***	33	1.2
V12.Pf3.3	NANPNVDP(NANP)$_3$-NVDPNANP SEQ ID NO:5	**	5	1
V12.PF3.4	NPNVDP(NANP)$_3$NV SEQ ID NO:6	****	5	5
V12.PF3.5	NPNVDP(NANP)$_3$NVDP SEQ ID NO:7	****	5	5
V12.PF3.6	NPNVDP(NANP)$_3$-NVDPNA SEQ ID NO:8	****	5	5
V12.PF3.7	NVDP(NANP)$_3$NV SEQ ID NO:9	-	-	-
V12.PF3.8	NVDP(NANP)$_3$NVDP SEQ ID NO:10	****	5	1
V12.PF3.9	NVDP(NANP)$_3$NVDPNA SEQ ID NO:11	***	5	ND
V12.PF3.10	DP(NANP)$_3$NV SEQ ID NO:12	****	1	ND
V12.PF3.11	DP(NANP)$_3$NVDP SEQ ID NO:13	****	5	ND
V12.PF3.12	DP(NANP)$_3$NVDPNA SEQ ID NO:14	***	5	ND

Immunogenicity of selected HBc chimer particles containing variants of the Pf3 epitope were assayed as described above.   Sera were analyzed by ELISA 4 weeks after primary (1°) and 4 weeks after booster (2°) immunizationsd.   The data obtained are shown in Table 6, below, in which the "Name" of the chimer and the corresponding sequence of the B cell immunogen are as illustrated above.

Table 6

NAME	PRIMARY	SECONDARY
V12.Pf1	40,960	655,360
V12.Pf3	2,621,440	10,485,760
V12.Pf3.1	2,621,440	10,485,760
V12.Pf3.2	2,621,440	2,621,440

Surprisingly, a version that contained one copy of the NANPNVDP repeat (V12.Pf3.1; SEQ ID NO: 126) was as immunogenic (and expressed better) as a version containing 2 copies (V12.Pf3), despite being 5-fold less antigenic for the NANP monoclonal antibody.

B. Expression failures

Several additional epitopes have been attempted to be placed into the HBc loop (Domain II) between positions 78 and 79 (as in V2.Pf1), and have failed to be expressed for reasons unknown.   Table 7,

-100-

below, enumerates those epitopes that have failed to express when inserted between D78 and P79 (V2) in a HBc chimer.

Table 7

Designation	Source of Epitope	Epitope (single letter)
V2.FGF-1 (N7-K12)	Human FGF-1	NYKKPK SEQ ID NO:127
V2.FGF-1 (K118-H124)	Human FGF-1	KRGPRTH SEQ ID NO:128
V2.Arom-479	P450 Aromatase	LHPDETKNMLEMIFTPRNSDR SEQ ID NO:129
V2.HIV3.1	HIV-1 (gp120)	RIKQI SEQ ID NO:130
V2.HIV4.1	HIV-1 (gp120)	RIKQIGMPGGK SEQ ID NO:131
V2.HIV5.1	HIV-1 (gp41)	LLELDKWASL SEQ ID NO:132
V2.HIV6.1	HIV-1 (gp41)	EQELLELDKWASLW SEQ ID NO:133
V2.HIV9.1	HIV-1 (gp41)	VQQQNNLLRAIEAQQHLL-QLTVWGIKQLQARIL SEQ ID NO:134
V2.HIV10.1	HIV-1 (gp41)	HLLQLTVWGIKQLQAR SEQ ID NO:135
V2.HIV12.1	HIV-1 (gp41)	YTHIIYSLIEQSQNQQEK-NEQELLALDKWASLWNWF SEQ ID NO:136
V2.HIV13.1	HIV-1 (gp41)	YTHIIYSLIEQSQN-QQEKNEQELLEL SEQ ID NO:137
V2.1A2(351-370)	Human P450-1A2	GRERRPRLSDRPQLPYLEA SEQ ID NO:138
V2.2D6(129-148)	Human P450-2D6	REQRRFSVSTLRNLGLGKKS SEQ ID NO:139
V2.Py-B1	P. yoelii (TRAP)	PNKLPRSTAVVHQLKRKH SEQ ID NO:140
V2.Py-B3	P. yoelii	TAVVHQLKRKH

	(TRAP)	SEQ ID NO:141
V2.Pv-T1A	P. vivax	PAGDRADGQPAGDRAAAGQPAG SEQ ID NO:142
V2.ALV1.2	ALV-J	NQSWTMVSPINV SEQ ID NO:143
V2.ALV1.2	ALV-J	MIKNGTKRTAVTFGSV SEQ ID NO:144
V2.FMDV (142-160)	FMDV	PNLRGDLQVLAQKVARTLP SEQ ID NO:145
V2.FMDV (135-160)	FMDV	RYNRNAVPNLRGDL- QVLAQKVARTLP SEQ ID NO:146

Example 5: <u>Determination of 280:260 Absorbance Ratios</u>

        Protein samples were diluted to a
concentration of between 0.1 and 0.3 mg/mL using
phosphate buffered saline (PBS), pH 7.4. The
spectrophotometer was blanked, using PBS, and the
absorbance of the protein sample measured at
wavelengths of 260 nm and 280 nm. The absorbance
value determined for a sample at 280 nm was then
divided by the absorbance value determined for the
same sample at 260 nm to achieve the 280:260
absorbance ratio for a given sample. The ratios
obtained for several samples, including native
particles (HBc 183), HBc particles truncated after
residue position 149 (HBc 149), and several HBc
chimers that are identified elsewhere herein, are
shown below in Table 8.

Table 8

280/260

Particle	Absorbance Ratio
HBc183	0.84
HBc149	1.59
V2.Pf1	1.64
V2.Pf1+C150	1.5
V2.Pf1 + Pf/CS-UTC	1.54
V2.Pf1 + Pf/CS-UTC(C17A)	1.42

Example 6: Cysteine at the C-terminus
          of Truncated HBc Particle

A. Addition of a Cysteine Residue
    to the C-terminus of Hybrid HBc Particles

Using the polymerase chain reaction (PCR),
genes expressing hybrid HBc particles can be easily
mutated to introduce a cysteine or cysteine-
containing peptide to the C-terminus of HBc. For
example, a PCR oligonucleotide primer such as that of
SEQ ID NO:148 can be used, in concert with a suitable
second primer, to amplify a hybrid HBc gene and
incorporate a cysteine codon between codon V149 and
the stop codon.

Hepatitis B core particles can be truncated
from residue position 183 (or 185, depending on viral
subtype) to 140 and retain the ability to assemble
into particulate virus-like particles. Many groups
have used particles truncated to amino acid 149

because amino acid 150 represents the first arginine residue of the arginine-rich C-terminal domain.

To assess the ability of a single cysteine residue to stabilize HBc particles, a codon for a cysteine residue was inserted using techniques described before between the codon for HBc amino acid residue V149 and the termination codon of a chimer HBc molecule that contained the $(NANP)_4$ malarial B cell epitope inserted between residues 78 and 79 (referred to herein as V2.Pf1) to form the chimeric molecule and particle referred to as V2.Pf1+C. The thermal stability (at $37^{\circ}C$) of this chimer particle (V2.Pf1+C), as compared to a similar chimer particle lacking the inserted cysteine (V2.Pf1), was found to be dramatically increased, as is seen in Fig. 3.

It is noted that vectors and expression products that are prepared by addition of a cysteine to the C-terminus of a V2 construct are sometimes referred to herein as V16 vectors or expression products.

As can readily be seen in Fig. 3, the two particles started out similarly. However, after fourteen days at $37^{\circ}C$, the cysteine-containing particle exhibited fewer bands on the SDS gel, indicating enhanced stability as compared to the particle lacking the added Cys residue.

B. Thermal Stability Protocol

Purified particles were diluted to a concentration of 1 mg/mL using 50 mM $NaPO_4$, pH 6.8 and sodium azide was added to a final concentration of 0.02% to prevent bacterial growth. Particles were incubated at 37° C and aliquots were taken at the time points indicated in the drawing description. Samples

were mixed with SDS-PAGE sample buffer (reducing) and
run on 15% SDS-PAGE gels. Gels were stained using
Coomassie Blue, and then analyzed.

Example 7:   Analytical Gel Filtration
             Analysis of Hybrid particles
          Analytical gel filtration analysis of
purified hybrid HBc particles was performed using a
25 mL Superose® 6 HR 10/30 chromatographic column
(Amersham Pharmacia # 17-0537-01) and a BioCAD™
SPRINT Perfusion Chromatography System. The UV
detector was set to monitor both wavelengths of 260
and 280 nm. The column was equilibrated with 3
column volumes (CV; about 75 mL) of buffer (50 mM
$NaPO_4$, pH 6.8) at a flow rate of 0.75 mL/minute.

          The particles to be analyzed were diluted
to a concentration of 1 mg/mL using 50 mM $NaPO_4$, pH
6.8. 200 Microliters (μL) of the sample were then
loaded onto a 200 μL loop and injected onto the
column. The sample was eluted from the column with
50 mM $NaPO_4$, pH 6.8 at a flow rate of 0.75 mL/minute.

          Particles containing C-terminal cysteine
residues or similar particles free of such cysteines
were analyzed using the above procedure. Integration
of the 280 nm trace was carried out using BioCAD™
software (PerSeptive™) to provide the results in
Table 9A, below.

Table 9A

Particle	Percent After Purification	
	Particulate	Non Particulate
V12.Pf1(C17A)	67	33
V12.Pf1(C17A) + C150 *	100	0
V12.Pf1 *	98	2
V2.Pf1+CfHBc74-87+C*	97.8	2.2
V2.Pf1+CfHBc74-87	80.7	19.3

* C-terminal cysteine-stabilized particles.

Purified particles were assayed for the percentage of particles and then incubated in aqueous solution at 37°C as discussed before. The compositions were assayed for stability after fourteen days of incubation. The results of this analysis are shown in Table 9B, below.

Table 9B

Particle Name	Percent Particles Following Incubations at 37°C (Days)	
	Zero	14
V12.Pf1 *	98	96
V12.Pf1(C17A)	67	63
V12.Pf1(C17A)+C150 *	100	98

* See the note to Table 9A.

Fig. 7 shows the results of a SDS-PAGE analysis of the particles of Table 9B at days zero, 7 and 14 following incubation at 37°C. Results of a densitometric analysis of that SDS-PAGE analysis are shown in Table 9C, below.

Table 9C

Particle	Percent Full Length Monomer Following Incubation at 37°C		
	Days		
	Zero	7	14
V12.Pf1 *	100	94	93
V12.Pf1(C17A)	100	13	1
V12.Pf1(C17A)+C150 *	100	83	63

* See the note to Table 9A.

The particles of Tables 9A-9C and control particles that contained an added lysine residue between HBc residues usually numbered 76 and 77 without [HBc150(K77)] and with a C-terminal Cys residue [HBc150(K77) + C] were analyzed for immunogenicity in BALB/c mice via intraperitoneal injection using 20 μg of the respective particles in phosphate buffered saline (pH 7.4) in the absence of adjuvant, contrary to the results reported in Example 4. Sera were analyzed two weeks after immunization using an ELISA with HBc particles (Anti-HBc) or (NANP)$_5$ synthetic peptide [Anti-(NANP)$_n$] as the solid phase capture antigen. The results of this study are shown in Table 9D, below

Table 9D

Particle	End Point Titer	
	Anti-HBc	Anti-(NANP)$_n$
V12.Pf1(C17A)	10,240	0
V12.Pf1 (C17A)+C150 *	10,240	2,560
V12.Pf1 *	10,240	10,240
HBc150(K77)	40,960	0

| HBc150(K77)+C* | 163,840 | 0 |

* See the note to Table 9A.

The data from this study are interpreted to mean that the C-terminal cysteine-stabilized particles are more stable immediately on production, as well as after incubation at 37°C for various time periods. The stabilized particles also exhibit enhanced immunogenicity, even in the absence of adjuvant. In addition, although particulate matter is present in the non-stabilized material such as V12.Pf1(C17A), there are no monomeric chimeric proteins after fourteen days of incubation and the material present does not induce antibodies toward the initially introduced heterologous B cell epitope sequence, here a malarial immunogen.

Example 8:   Cysteine Located Within a Peptide
            Fused to the C-terminus of an HBc Hybrid
            Studies were conducted to determine if there were an absolute requirement for a cysteine residue to be the final amino acid of the HBc gene (as it is in wild type HBc) or if a cysteine could function internally in an introduced C-terminal sequence.

A peptide corresponding to a 20-residue universal T cell epitope, derived from the CS protein of the malarial parasite *Plasmodium falciparum*, which contains a cysteine at position 17 of the peptide or 342 of the CS protein, [Calvo-Calle et al., *J. Immunol.*, (1997) 159(3):p. 1362-1373], was fused to the C-terminus of a HBc chimer (V2.Pf1). This chimer contains the HBc sequence from position 1 through

position 149, with the *P. falciparum* B cell epitope
(NANP)$_4$ inserted between amino acid residues 78 and
79. Domain I of this HBc construct thus contained
residues 1-75; Domain II contained residues 76-85
with the (NANP)$_4$ epitope inserted between residues 78
and 79 (along with four residues comprising the
restriction sites); Domain III contained residues 86-
135; and Domain IV contained residues 136-149 plus
the 20-residue *P. falciparum* T cell epitope and two
residues from the EcoRI cloning site (GI).

This fused C-terminal peptide is 20 amino
acid residues long (12 or 14 amino acids shorter than
the wild type sequence, depending on virus subtype)
and has a predicted pI value more than 8 pH units
lower than the wild type sequence. To minimize
potential stabilizing effects that may be contributed
by amino acids other than the cysteine, a (similar)
control construct was made, having an alanine instead
of a cysteine at position 17 (see Table 10, below).

To enable simple assessment of the
stabilizing effects of this sequence, the peptides
were fused to the C-terminus of a particle previously
shown to degrade readily at 37°C (V2.Pf1) to form the
HBc chimers denominated V2.Pf1+Pf/CS-UTC and
V2.Pf1+Pf/CS(326-345/C342A), respectively. The
results of a thermal stability study over a 28 day
time period (as discussed previously) are shown in
Fig. 4.

The results of this study showed that the
presence of the cysteine in the T cell epitope
derived from the CS protein of *P. falciparum* was
needed for particle stability in the time period
studied, and that there was no absolute requirement
that that cysteine be at the C-terminus of the

epitope. The table below shows the amino acid
sequences of C-terminal fusions with a cysteine or
alanine at position 17, relative to the native
sequence, which occurs in the wild type HBc protein.

Table 10

Source	Sequence	pI	Length	Cys Position	Cys Shift
Native	RRRGRSPRRRT-PSPRRRRSQSP-RRRRSQSRESQC SEQ ID NO:147	12.74	34	34	zero
Pf/CS-UTC	(GI)EYLNKIQNS-LSTEWSPCSVT SEQ ID NO:148	4.44	20	17	-15
Pf/CS-UTC(C17A)	(GI)EYLNKIQNS-LSTEWSPASVT SEQ ID NO:149	4.44	20	N/A	N/A

(GI) = residues added from cloning site.

Example 9: *P. Vivax* HBc Chimers

        Following the work discussed before on HBc
chimers containing *P. falciparum* B cell and T cell
immunogens, similar work was carried out using
sequences from the *P. vivax* CS protein. Exemplary
constructs are illustrated below in Table 11.

Table 11

P. vivax Immunogen Type	Malarial B Cell Immunogen (Between D78/P79)	CS-UTC (After V149)
Type-I	(DRAAGQPAG) SEQ ID NO:152 (DRADGQPAG) SEQ ID NO:153	YLDKVRATVGTEWTPCSVT SEQ ID NO:25
Type-II	(ANGAGNQPG) SEQ ID NO:154 (ANGAGDQPG) SEQ ID NO:155 (ANGADNQPG) SEQ ID NO:156 (ANGADDQPG) SEQ ID NO:157	YLDKVRATVGTEWTPCSVT SEQ ID NO:25
Type-III ('Vivax-like')	(APGANQEGGAA) SEQ ID NO:158	YLDKVRATVGTEWTPCSVT SEQ ID NO:25

To address the variability of the repeats, the following variant epitopes were used for insertion into HBc between amino acids 78 and 79:

1.   Type-I CS-repeat

PAGDRADGQPAGDRAAGQPAG (*P. vivax*-type 1A)--SEQ ID NO: 159. This form of the epitope <u>failed</u> to make a particle.

DRAAGQPAGDRADGQPAG (*P. vivax*-type 1B)-- SEQ ID NO: 150. This form of the epitope, containing flanking dipeptide cloning site remnants, successfully made a particle and is referred to as V2.PV-TIB. An immunogen for *P. vivax*-type I has been successfully cloned, expressed, purified, and its immunogenicity

-111-

tested in mice. The results of that mouse study are shown in Table 12, hereinafter.

2. Type-II CS-repeat

For type-II, this work is complicated by the existence of four different forms of the type-II epitope. These forms contain either G or D at position 5, and either N or D at position 6 [Qari et al., *Mol. Biochem. Parasitol.*, (1992) 55(1-2):p. 105-113]. Hence, there are 4 different possible repeat sequences (GN, GD, DN, and DD) needed to maximize the possibility of success. The first, and preferred approach, is to prepare a single hybrid particle containing all four repeats, as shown below by underlines. This approach was successfully employed to address the variability in the type-I repeat. Each of these constructs contains flanking dipeptide cloning site remnants.

ANGA<u>GN</u>QPGANGA<u>GD</u>QPGANGA<u>DN</u>QPGANGA<u>DD</u>QPG

(*P. vivax*-type II -GN/GD/DN/DD)   SEQ ID NO: 151.

The above sequence has been cloned, expressed, and purified as a HBc chimer with no modification to the C-terminus.

The second approach was to prepare two hybrid particles, whereby each particle contained two of the variant epitopes (see below). This approach is less preferable because it requires either the use of a more complex expression system to direct the production of 'mixed' particles during expression, or the mixing of type-II particles following manufacture.

ANGA<u>GN</u>QPGANGA<u>GD</u>QPG   (*P. vivax*-type II-GN/GD)
SEQ ID NO: 160.

QANGA<u>DN</u>QPGANGA<u>DD</u>QPG   (*P. vivax*-type II-DN/DD)
SEQ ID NO: 161.

CGCGAATTCAAGCGAACGGCGCCGATAATCAGCCGGCGGGTGCA
(*P. vivax*-type IIB-ER1-wt-F)   SEQ ID NO: 162.

3.   Type-III ('vivax-like') CS-repeat

The third *P. vivax* CS-epitope, which is quite
different from the other two, is not associated with
amino acid variation (see below) [Qari et al.,
Lancet, 1993. 341(8848): p. 780-783]. This sequence
was cloned into the HBc expression system, and
hybrids were produced that contained flanking
dipeptide cloning site remnants.

APGANQEGGAAAPGANQEGGAA (*P.vivax*-type III)
SEQ ID NO: 163.

4.   T cell Epitope at the C-terminus of HBc

The insertion of the *P. vivax* Th epitope
(Pv-UTC; YLDKVRATVGTEWTPCSVT; SEQ ID NO:25) into HBc
and HBc hybrids was also performed using synthetic
DNA fragments (Synthetic Genetics, San Diego CA).
However, unlike B cell epitopes, which are inserted
into the immunodominant loop region of the HBc gene,
T cell epitopes are fused to the C-terminus of the
HBc gene. Previously discussed cloning vectors were
used for the insertion of both B and Th epitopes into
HBc. The particle expressing just the Pv-UTC at the
C-terminus has also been successfully made.

5. Combining B and T cell Epitopes
in a Single Particle

To combine B and Th epitopes into single HBc constructs, PCR is used to amplify N-terminal HBc fragments (residues 1-80, which contain the B cell epitopes), and C-terminal HBc fragments (residues 81-150, which contain the T cell epitopes). The fragments are ligated together and amplified again by PCR. Again, clones are verified by restriction endonuclease mapping and automated DNA sequence analysis (Lark Technologies, Houston TX). Details are essentially the same as for *P. falciparum*. Particles that contain each of the Type-I, -II and -III B cell epitopes and variants as well as the Pv-UTC, have been expressed and recovered.

Example 10: Relative Immunogenicities of HBc Chimers

Relative immunogenicities of several HBc chimer immunogens were compared in mice using the IFA assay discussed previously. The results of those studies using two dose immunization regimens as before are shown below in Table 12.

Table 12

Immunogen	IFA titer	Protection	Citation
P. berghei (CS-1)	40,960	95%	A
P. yoelii (CS-3)	12,800	95%*	B
P. falciparum (CS-2)	1,200	NT	A
P. falciparum	5,200,000	NT	--

(V12.Pf3.1)

P.vivax (V2.PV-TIB)   160,000   NT   --

[A = Schodel et al., J. Exp. Med., 1994, 180:1037-1046. B
= Schodel et al., *Behring Inst. Mitt.*, 1997(98): p. 114-
119. NT = not tested. * = protection for greater than 3
months.]

    As is seen from the above data, titers of
10^5-10^6 for *P. falciparum* were achieved using a
contemplated chimeric immunogen; this compares to
titers of only 10^4 for *P. berghei* and 10^3 for *P.
falciparum* using the replacement technology of
Schodel et al.

    Mice were immunized with CS-2 or V12.Pf1
using 20 µg of particles on day zero and were boosted
with 10 µg at four weeks. Mice immunized with
particles from V12.Pf3 and V12.Pf3.1 were immunized
using 20 µg of particles on day zero and were boosted
with 10 µg at eight weeks using adjuvants as
discussed before. Data showing the duration of the
titers achieved are shown in Fig. 5, with data for
use of V12.Pf3 particles being essentially identical
to data with V12.Pf3.1 particles, and not shown.

Example 11: <u>Relative HBc antigenicities</u>

    A series of studies was carried out to
determine the relative antigenicities of several
malarial HBc chimer particles toward two monoclonal
antibodies (MoAb-3120 and MoAb-3105) as compared to
native HBcAg (particle). These antibodies are
specific to the loop region of HBc, and were the

gracious gift of the Immunology Institute, Tokyo,
Japan. Studies were carried out using the chimers of
Table 5 that contain malarial epitopes inserted into
HBc particles at various positions as antigens in
ELISA assays with the monoclonals as probes. The
results of these studies (as end point dilutions) are
shown below in Table 13A, 13B, and 13C, and
illustrate the substantial lack of antigenicity of a
contemplated chimer toward monoclonal antibodies that
bind to the loop region, the primary immunogen, of
HBc. Put differently, monoclonal antibodies that
bind specifically to the loop region of HBc barely
recognize a contemplated chimer, if at all.

Table 13A

Particle	Anti-MoAb-3120 End Point Dilution	Relative Antigenicity
HBcAg	625000	100
V12.Pf3	80000	12.8
V12.Pf3.1	20000	3.2
V12.Pf3.2	10000	1.6
V12.Pf3.3	10000	1.6
V12.Pf3.4	80000	12.8
V12.Pf3.5	40000	6.4
V12.Pf3.6	80000	12.8
V12.Pf3.8	80000	12.8
V12.Pf3.9	160000	25.6
V12.Pf3.10	10000	1.6
V12.Pf3.11	80000	12.8
V12.Pf3.12	80000	12.8

Table 13B

Particle	Anti-MoAb-3105 End Point Dilution
HBcAg	1,300,000
V2.Pf1 (78/79)	Zero
V12.Pf1 (78/79)	Zero
V12.Pf3 (78/79)	Zero
V1.Pf1 (77/78)	Zero
V13.Pf1	1,300,000

An insertion into several sites in the
immunodominant loop (including positions 77-78 or 78-
79) totally eliminates binding of MoAb-3105. V13 is
an insertion between residues 129 and 130 and is used
as a control because the native HBc immunodominant
loop remains intact in this construct.

Table 13C

Particle	Anti-MoAb-3120 End Point Dilution
77/78 V1.Pf1	102,400
78/79 V2.Pf1	400
HBcAg	409,600

These data show that insertion between
residues 78 and 79 causes a more drastic reduction in
anti-MoAb-3120 binding as compared with insertion
between residues 77 and 78.

Example 12:   Preparation of Vector for Preparation
              of HBc Particles for Use in Humans

A. Preparation of Vector V17Pf3.1

To manufacture the particle V12.Pf3.1 in a
manner suitable for human administration, it was
necessary to express the particle using an expression
system that did not require the use of ampicillin to
ensure plasmid maintenance. To achieve this, the
gene coding for the particle, along with the
necessary upstream regulatory sequences, was inserted
into a new plasmid that utilizes kanamycin as the
selectable marker. The new plasmid (V17.Pf3.1) was
synthesized using a two step cloning procedure:

Step 1: The plasmid pKK223-3N-V12 was
digested with the restriction enzymes BamHI and
HindIII to yield two DNA fragments of 801 and 4869
bp. In addition, the commercially available plasmid
pREP4 (Qiagen) was cut with BglII and HindIII to
yield two fragments of 320 bp and 3420 bp. The 3420
bp and 801 bp fragments were ligated to create
plasmid V17. (It is noted that BglII and BamHI
digested DNAs can be ligated by virtue of their
common 'overhang' sequences, although neither BglII
or BamHI can cut the resultant fragment). The V17
plasmid, therefore, contains the HBc149 gene,
complete with Pf-UTC sequence fused to the C-
terminus, and EcoRI and SacI restriction sites in the
immunodominant loop region to enable insertion of
epitopes between D78 and P79 of the HBc gene.

Step 2: The second step was to insert the
Pf3.1 version of the Pf CS-repeat epitope into the
immunodominant loop region of the gene. This was
achieved by digesting V17 with SacI and EcoRI to

yield 15 bp and 4206 bp DNA fragments. Annealed
oligonucleotides encoding the Pf3.1 epitope were
ligated with the 4206 bp fragment to yield V17.Pf3.1,
a 4275 base pair plasmid. In addition to the gene
that encodes the 195 amino acid malaria vaccine
candidate, this plasmid contains a gene for the lac
repressor (lac I) to force any gene under lac
promoter control to be fully repressed until induced
by isopropylthiogalactoside (IPTG). It also has a
kanamycin resistance gene to permit positive
selection via the addition of kanamycin to culture
media. The plasmid has the replication origin of
pACYC 184 and is not considered to be a high copy
number plasmid.

The locations of the genes of interest are:

Gene	Start	Stop	Amino Acids	Molecular Weight (kDa)
Lac I	2128	3087	319	34.1
V17.Pf3.1	281	868	195	21.7
KmR	4259	3465	264	29.1

A suitable host for V17.Pf3.1 is *E. coli*
BLR, a rec A derivative of *E.coli* BL21, and a common
strain used for the production of recombinant
proteins (available for purchase from Novagen). *E.
coli* BLR was selected as a host organism for
expression because of its increased genetic
stability, as well as its ability to produce

assembled particles in soluble form (not in inclusion bodies).

B. Expression of Particles

Using Plasmid V17.Pf3.1

E.coli (Strain BLR) containing the V17.Pf3.1 plasmid were streaked onto an LB agar plate supplemented with 25 μg/mL kanamycin and 10 μg/mL tetracycline, then incubated at 37°C for 16-20 hours. A single colony was then used to inoculate 3 mL of TB-Phy medium in a sterile culture tube, supplemented with 25 μg/mL kanamycin. The tube was incubated overnight (about 18 hours) on a shaker at 37°C and about 200 rpm.

The following morning, 100 mL of TB-Phy medium was warmed to 37°C. One mL of the overnight culture was removed and used to inoculate the flask, which was then incubated on a shaker at 37°C at about 200 rpm for six hours.

The fermentor (Biostat™ UE20) was inoculated with 100 mL of inoculum with the fermentor conditions set as follows:

Agitation	400 rpm
Temperature	37°C
Aeration	air, 10 liters per minute
pH	7.0, uncontrolled

The A_{600} value was measured for the first sample, and for samples every 20-30 minutes thereafter to monitor A_{600}. An IPTG solution was prepared by dissolving 62 mg IPTG in 10-15 mL water. When the A_{600} value reached 0.5, the filter-

sterilized IPTG solution was aseptically added to the fermentor through a syringe. The incubation was continued until next day (e.g. about another 10-24 hours).

At 14 hours after induction, the fermentor temperature was set to 15°C. Harvesting of cells was started by centrifugation in a Beckman® J2-MC centrifuge with following conditions:

Rotor	JA10
Speed	7,500 rpm
Temperature	4°C
Time	9 minutes

The cells were harvested by freezing into liquid nitrogen.

C.   Purification of Particles
     Expressed by Vector V17.Pf3.1/BLR

The biomass of harvested cells was resuspended in 50 mM sodium phosphate, pH 6.8, and lysed using a French Pressure cell at 16,000 psi. The cell debris was removed by centrifugation using a Beckman® J2-MC centrifuge and the following conditions.

Rotor:	JA20
Speed:	15,000 rpm
Temperature:	4°C
Time:	30 minutes.

The volume of the resultant supernatant was measured and 277 g/L of solid ammonium sulfate were slowly added to the supernatant. The mixture was

-121-

stirred at 4°C for 30 minutes. The solution was centrifuged in Beckman® J2-MC centrifuge with the following conditions.

Rotor: JA20
Speed: 15,000 rpm
Temperature: 4°C
Time: 30 minutes

The precipitate was then resuspended in a minimal volume of 50 mM sodium phosphate buffer and then dialyzed against the same buffer for one hour with stirring. The dialyzed solution was centrifuged in Beckman® J2-MC centrifuge with the following conditions.

Rotor: JA20
Speed: 15,000 rpm
Temperature: 4°C
Time: 15 minutes

The supernatant was recovered and then subjected to gel filtration chromatography.

System: Pharmacia Biotech AKTA™ Explorer
Buffer B (elution solvent): 50 mM Sodium
            phosphate buffer (pH 6.8).
Column: Millipore Vantage™ VL44 x 1000 column (44 mm
            diameter, 1000 mm height, Catalog No.:
            96441000)
Resin: 1.5 liter Sepharose® CL-4B manufactured by
            Pharmacia
Detector: UV at 210, 254 and 280 nm.

Fraction: 15 mL

        The column was eluted with buffer B at 2 mL
per minute. Particle-containing fractions were
identified using SDS-PAGE and pooled. The salt
concentration of the pooled material was adjusted to
5M by adding sodium chloride.

Hydrophobic Interaction Chromatography:

System: Pharmacia® Biotech AKTA™ Explorer
         (System No.: 18111241 001152,
         University of Iowa ID No.: 540833.)

Buffer A: 50 mM sodium phosphate buffer
           (pH 6.8) + 5 M NaCl. (The buffer
           was degassed for 30 minutes
           daily, before use.)

Buffer B (elution solvent): 50 mM sodium
           phosphate buffer (pH 6.8). (The
           buffer was degassed for 30 minutes
           daily, before use.)

Hydrophobic Interaction Chromatography using
ToyoPearl® ether 650 resin

Column: Millipore Vantage™ VL44 x 250
          column (44 mm diameter, 250 mm
          height, Catalog No.: 96440250)
Resin: 200 mL Toyopearl® ether 650 HIC
          resin, manufactured by Tosohaas
Detector: UV at 210, 254, and 280 nm

Fraction: 15 mL

      The column was equilibrated with 5 column
volumes (CV) of buffer A for a one hour time prior to
starting purification, using a flow rate of 20
mL/minute. The retentate containing 5 M salt was
then loaded at a rate of 20 mL/minute. The column
was washed with 2 CV of buffer A, washed with 2 CV of
10% buffer B, eluted with 3 CV of 40% buffer B, and
(finally eluted) with 100 % buffer B. Fractions were
completely analyzed for proteins of interest by SDS
PAGE analysis. Pure fractions were combined
together, and a protein estimation using a Bradford
assay was carried out.

Hydrophobic Interaction Chromatography using butyl
resin

Column: Millipore Vantage™ VL44 x 250
        column (44 mm diameter, 250 mm
        height, Catalog No.: 96440250)
Resin: 200 mL Toyopearl® Butyl 650-S HIC
        resin, manufactured by Tosohaas
Detector: UV at 210, 254 and 280 nm
Fraction: 15 mL

      The column was equilibrated with 5 column
volumes (CV) of 40% buffer B for one hour prior to
starting purification, using a flow rate of 20
ml/min. The combined fractions from ether HIC were
loaded at a rate of 20 mL/minute. The column was
washed with 2 CV of 40% buffer B, washed with 2 CV
90% B, and eluted with 4 CV of WFI.

Fractions were analyzed for protein of interest by SDS PAGE analysis. Pure fractions were combined together

Hydroxyapatite Column Chromatography

Column: Millipore Vantage™ VL16 x 250
          column (16 mm diameter, 250 mm
          height, Catalog No.: 96160250)
Resin: 20ml Ceramic Hydroxyapatite (Catalog No.
          158-2200)
Detector: UV at 215, 254 and 280 nm
Fraction: 15 mL

          The column was equilibrated with 5 column volumes (CV) of 20 mM sodium phosphate buffer, flow rate: 5 mL/min. Load combined fractions eluted from butyl HIC at 5 mL/min. Wash the column with 20 mM sodium phosphate buffer until A280 drops to baseline. Fractions were analyzed for protein of interest by SDS PAGE analysis. Pure fractions were combined together.

Desalting

Column: Prepacked desalting column, HiPrep™ 26/10,
          Pharmacia
Resin: 20 mL Ceramic Hydroxyapatite (Catalog No.
          158-2200)
Detector: UV at 215, 254 and 280 nm
Fraction: 15 mL

The column was equilibrated with 5 CV of 15
mM Acetate Buffer, pH 6.0. The pooled fractions from
the hydroxyapatite column were loaded onto the
column, and then eluted with 15 mM Acetate Buffer, pH
6.0, at a flow rate of 20 mL/min. Fractions were
analyzed for protein of interest by SDS PAGE
analysis. Pure fractions were combined together, and
protein estimation was carried out using a Bradford
assay. The pure fraction was assayed for endotoxin
level, and finally passed through a 0.22-micron
filter for terminal filtration.

Example 13:   Preparation of Vectors to Express
              Particles with a Cysteine Residue Prior
              to C-Terminal Fused Epitope

          To prepare particles with a single cysteine
after V149 of the HBc gene, followed by a T cell
epitope, a PCR primer was synthesized (SEQ ID
NO:165). This primer, in conjunction with
HBc149/NcoI-F (SEQ ID No:29), was used to amplify the
HBc gene to produce a version of HBc having a single
cysteine codon introduced directly after V149, as
well as EcoRI and HindIII restriction sites (after
the introduced cysteine). The 478 bp PCR product was
cut with NcoI and HindIII and cloned into pKK223-3N.

                                  SEQ ID No.: 164

                         C   V   V   T   T   E   P

5' GCAAGCTTACTATTGAATTCCGCAAACAACAGTAGTCTCCGG
       HindIII   EcoRI
                                  SEQ ID NO:165

The resultant plasmid was then cut with
EcoRI and HindIII, and the annealed oligonucleotides
coding for the Pf-UTC (Pf/CS326-345; SEQ ID NOs: 80
and 81) ligated into the plasmid. This plasmid was
then used as the template in a PCR reaction, along
with the primers HBc-P79/SacI-F (SEQ ID NO:35) and
Pf/CS-UTC(C17A) (SEQ ID NO:104) the resultant PCR
product (307 bp) coded for amino acids 79 through 149
of HBc, followed by the introduced cysteine, followed
by the Pf/CS-UTC having the C17A mutation [Pf/CS-
UTC(C17A)], and flanked by SacI (5') and HindIII (3')
restriction sites. This fragment was cut with SacI
and HindIII and ligated with the plasmid V2.Pf1
(encoding the malarial (NANP)$_4$ epitope) that had been
cut with the same two enzymes.

The resultant gene codes for a 190 amino
acid HBc chimera having (NANP)$_4$ inserted between
amino acids 78 and 79 of HBc, (flanked by the Gly-Ile
and Glu-Leu sequences derived from the EcoRI and SacI
restriction sites respectively) and the C17A version
of the Pf/CS-UTC at the C terminus. The single
cysteine was therefore located between V149 of HBc
and the Gly-Ile linker sequence (derived from the
EcoRI restriction site) located prior to the first
amino acids of the Pf/CS-UTC(C17A) T cell epitope
(see SEQ ID NO:167).

This hybrid particle was expressed, purified and
analyzed for stability by incubating at 37°C for
several weeks. The stability of this particle
(V12.Pf1(C17A)C150) was compared to V12.Pf1, with the
only difference between the two particles being the
position of the cysteine residue. For V12.Pf1 the
cysteine is followed by 3 amino acids (SVT) at the C-
terminus of the protein (SEQ ID No: 166), whereas for

V12.Pf1(C17A)C150 the cysteine is followed by 22
additional amino acid residues (SEQ ID No: 167).

<u>V12.Pf1</u>

   TTVV   GI EYLNKIQNSLSTEWSP<u>C</u>SVT   SEQ ID No:166

<u>V12.Pf1(C17A)C150</u>

   TTVV <u>C</u> GI EYLNKIQNSLSTEWSP<u>A</u>SVT   SEQ ID No:167

     The effect of inserting the cysteine
residue between HBc and the T cell epitope
(V12.Pf1(C17A)C150) was to create a particle that was
significantly more stable than a similar particle
without the C terminal cysteine (V12.Pf1(C17A)).
This was evident from the fact that unlike
V12.Pf1(C17A), V12.Pf1(C17A)C150 could be easily
purified without a significant degree of degradation
of monomers (compare T=0 for these particles in Figs.
.4 and 7); and further, V12.Pf1(C17A)C150 was
significantly more stable than V12.Pf1(C17A)

following incubation at 37°C. After 14 days at 37°C,
V12.Pf1(C17A) monomers are totally degraded (Fig. 4),
whereas V12.Pf1(C17A)C150 monomers are only partially
degraded (Fig. 7).

     It was apparent that V12.Pf1(C17A)C150 was
not as stable V12.Pf1 (Fig. 7). These data indicate
that the stabilizing effects of a single C-terminal
cysteine residue are most effective when placed at or
near, e.g. within five residues of, the C-terminus of
the HBc chimer.

Example 14:   Preparation of Vectors to Express
              Particles with a Cysteine Residue at
              <u>the C-Terminus of a Fused Epitope</u>

                         .

                       -128-

To further investigate whether terminal
cysteine residues could elicit stabilizing effects at
positions other than 150, a Th epitope from the
hepatitis B core protein (amino acid residues 74-87)
was fused to the C-terminus of HBc containing a
malarial epitope [(NANP)$_4$] in the immunodominant
loop. This Th epitope does not contain a cysteine
residue, so a Cys residue was added at the C-terminus
(underlined "C"). The control contained the same
fused epitope lacking the cysteine.

These particles were made by combining
V2.Pf1 with V7.HBc74-87 (and V7.HBc74-87+C) to form
the desired vector. The V7 construct was PCR
amplified with the HBc-P79/SacI-F primer (SEQ ID NO:
35) and pKK223-2/4515-32-R (SEQ ID NO:36). The
product was cut with SacI and HindIII, and the
SacI/HindIII fragment was ligated into V2.Pf1 cut
with the same enzymes.

Table 14, below, shows the amino acid
sequences of C-terminal fusions HBc(74-87) and
HBc(74-87) + C, relative to the native sequence that
occurs in the wild type HBc protein, as well as the
and the HBc149 + C particle. "Cys shift" is the
position of the introduced cysteine relative to its
location in the wild type protein, where it is the
last residue (position 183).

Table 14

Source	Sequence	PI	Length	Cys Position	Cys Shift
Native	RRRGRSPRRRT- PSPRRRRSQSP- RRRRSQSRESQC SEQ ID NO:147	12.74	34	34	Zero
HBc(74-87)	GIVNLEDPAS- RDLVVS SEQ ID NO:182	3.78	16	N/A	N/A
HBc(74-87)+C	GIVNLEDPAS- RDLVVSC SEQ ID NO:183	3.78	16	16	-17
HBc-149+C	C	N/A	1	1	-33

Example 15: Comparative Immunogenicities in Monkeys

The comparative immunogenicity of the particles expressed by V12.Pf3.1, formulated with either Seppic™ ISA-720 (Seppic Inc., Paris, France), Alhydrogel™ (Superfos, Denmark) as adjuvants, or unformulated (saline), was studied in Cynomolgus monkeys.

The Seppic™ ISA-720 formulation was prepared according to the manufacturers directions. Briefly, the ISA-720 and V12.Pf3.1 particles were mixed at 70:30 (w/w) ratio and vortexed, using a bench top vortexer, set at maximum power, for 1 minute. The Alhydrogel™ formulation was prepared using an 8-fold excess of Alhydrogel™ (by weight) over V12.Pf3.1 particles, which was shown to be physically bound to the Alhydrogel™ prior to immunization.

Groups of two monkeys (one male and one female) were immunized with 20 μg V12.Pf3.1 particles as immunogen via the intramuscular route. Animals were bled on days 0, 21, 42, 56 and 70, and sera analyzed for titers of anti-NANP antibody using an ELISA. The results, shown in Table 15, below,

demonstrate the extremely high immunogenicity of
V12.Pf3.1 particles when formulated with Seppic™ ISA-
720 versus Alhydrogel™-formulated or unformulated
material. The kinetics of the antibody response were
more rapid when Seppic™ ISA-720 was used as the
adjuvant, and the end-point titers were more than
100- and 1000-fold higher than for Alhydrogel™ and
saline respectively.

Table 15

Adjuvant	Antibody Titers at Stated Time (Days)				
	Zero	21	42	56	70
Saline	Zero	40	240	1,200	640
Anhydrogel™	Zero	2,880	1920	11,500	6400
Seppic™ ISA-720	Zero	81,920	348,160	26,000,000	1,920,000

Example 16: T Cell Activation

Mice were immunized twice with V12.Pf3.1
particles in Seppic™ Montanide™ ISA-720. Spleen
cells were removed and stimulated in the presence of
various peptides. 10^6 cells were incubated for 3 days
in the presence of peptides: UTC (universal T epitope
from P. falciparum; Seq IN NO: 120), p85-100 peptide
corresponding to HBc 85-100, NANP (B-cell epitope
from V12.Pf3.1; NANPNVDP(NANP)$_3$ SEQ ID NO:22) in the
presence of Staphylococcal enterotoxin B (SEB), or
tissue culture medium (unstim). Interferon gamma
production after 3 days was determined by ELISA.

The results shown in Table 16, below,
indicate that immunizing with V12.Pf3.1 induces T-
cells that recognize the UTC component of the
protein, and drives them to a Th1 type response.

Table 16

Immunogen	IFN-γ (pg/ml)	S.D.*
UTC	1600	750
p85-100	350	30
NANPNVDP (NANP)3 SEQ ID NO:22	370	50
SEB	4300	ND**
unstim	900	1100

* S.D. = Standard Deviation
** ND = Not Done

Each of the patents and articles cited
herein is incorporated by reference. The use of the
article "a" or "an" is intended to include one or
more.

The foregoing description and the examples
are intended as illustrative and are not to be taken
as limiting. Still other variations within the
spirit and scope of this invention are possible and
will readily present themselves to those skilled in
the art.

What is Claimed:

        1. A recombinant hepatitis B virus core
(HBc) protein chimer molecule with a length of about
140 to about 310 amino acid residues that contains
four peptide-linked amino acid residue sequence
domains from the N-terminus that are denominated
Domains I, II, III and IV, wherein

        (a) Domain I comprises about 71 to about 85
amino acid residues whose sequence includes at least
the sequence of the residues of position 5 through
position 75 of HBc;

        (b) Domain II comprises about 18 to about
58 amino acid residues peptide-bonded to residue 75
of which (i) a sequence of HBc is present from HBc
positions 76 through 85 and (ii) a sequence of 8 to
about 48 residues that constitute a B cell epitope of
the circumsporozoite protein of a species of the
parasite *Plasmodium* that is peptide-bonded between
the HBc residues of positions 78 and 79;

        (c) Domain III is an HBc sequence from
position 86 through position 135 peptide-bonded to
residue 85; and

        d) Domain IV comprises (i) zero to fourteen
residues of a HBc amino acid residue sequence from
position 136 through 149 peptide-bonded to the
residue of position 135 of Domain III,, (ii) zero to
three cysteine residues, (iii) fewer than three
arginine or lysine residues, or mixtures thereof
adjacent to each other, and (iv) up to 100 amino acid
residues in a sequence heterologous to HBc from
position 150 to the C-terminus, with the proviso that
at least five amino acid residues are present of the

amino acid residue sequence from position 136 through
149, when (a) zero cysteine residues are present and
(b) fewer than about five heterologous amino acid
residues are present.

2. The recombinant HBc chimer protein
molecule according to claim 1 present as self-
assembled particles.

3. The recombinant HBc chimer protein
molecule according to claim 1 wherein said B cell
epitope comprises two to about five repeats of a 4 to
about 11 amino acid residue sequence.

4. The recombinant HBc chimer protein
molecule according to claim 3 wherein said *Plasmodium*
species is *falciparum*.

5. The recombinant HBc chimer protein
molecule according to claim 3 wherein said *Plasmodium*
species is *vivax*.

6. The recombinant HBc chimer protein
molecule according to claim 1 wherein Domain I
consists essentially of the HBc sequence from
position 1 through position 75.

7. The recombinant HBc chimer protein
molecule according to claim 1 wherein Domain II
independently includes zero to three peptide-bonded
residues on either side of said B cell epitope that
are other than those of HBc or said B cell epitope.

8. The recombinant HBc chimer protein molecule according to claim 1 wherein said sequence heterologous to HBc at position 150 to the C-terminus of Domain IV comprises an amino acid residue sequence that constitutes a T cell epitope of the same species of *Plasmodium* as said B cell epitope.

9. A recombinant hepatitis B virus core (HBc) protein chimer molecule with a sequence of about 155 to about 235 amino acid residues that contains four peptide-linked domains from the N-terminus that are denominated Domains I, II, III and IV, wherein

(a) Domain I consists essentially of a sequence of residues 1 through 75 of HBc;

(b) Domain II is about 18 to about 46 residues in length of which (i) 10 residues are present in a sequence of HBc at positions 76 to 85 and (ii) a sequence of 8 to about 36 residues that constitute a B cell epitope of the circumsporozoite (CS) protein of *Plasmodium falciparum* or *Plasmodium vivax* that is peptide-bonded between the residues of positions 78 and 79, said B cell epitope being comprised of two to about five repeats of an amino acid residue sequence, said Domain independently including zero to three peptide-bonded residues on either side of said B cell epitope that are other than those of HBc or said B cell epitope;

(c) Domain III consists essentially of the HBc sequence from position 86 through position 135; and

d) Domain IV comprises a sequence of HBc from residue 136 through 140 peptide-bonded to the residue of position 135 of Domain III and (i) zero to

nine residues of a HBc amino acid residue sequence
from position 141 through 149, (ii) zero to three
cysteine residues, (iii) fewer than three arginine or
lysine residues, or mixtures thereof adjacent to each
other, and (iv) up to 50 amino acid residues in a
sequence that constitutes a T cell epitope of the
same species of *Plasmodium* as said B cell epitope
peptide-bonded to the final HBc amino acid residue
present in the chimer.

10. The recombinant HBc chimer protein
molecule according to claim 9 wherein Domain IV
comprises one amino acid residue to a sequence of
about nine amino acid residues of the HBc sequence
from residue position 141 through about position 149
peptide-bonded to residue 140.

11. The recombinant HBc chimer protein
molecule according to claim 10 wherein Domain IV
consists essentially of a sequence of nine amino acid
residues of the HBc sequence from residue position
141 through position 149 peptide-bonded to residue
140.

12. The recombinant HBc chimer protein
molecule according to claim 9 wherein the repeated
sequence of said B cell epitope of Domain II is SEQ
ID NO:152 and SEQ ID NO:153.

13. The recombinant HBc chimer protein
molecule according to claim 9 wherein the repeated
sequence of said B cell epitope of Domain II is Asn-
Ala-Asn-Pro.

14.   The recombinant HBc chimer protein molecule according to claim 13 wherein the repeated sequence of said B cell epitope of Domain II is repeated three or four times.

15.   The recombinant HBc chimer protein molecule according to claim 14 wherein the repeated sequences are peptide-bonded to each other without interruption.

16.   The recombinant HBc chimer protein molecule according to claim 15 wherein said B cell epitope includes a second CS protein sequence from the same *Plasmodium* species that is peptide-bonded to said repeated sequence.

17.   The recombinant HBc chimer protein molecule according to claim 16 wherein said second CS protein sequence is Asn-Val-Asp-Pro.

18.   The recombinant HBc chimer protein molecule according to claim 17 wherein said second CS protein sequence is peptide-bonded at the carboxy-terminus of said repeated sequence.

19.   The recombinant HBc chimer protein molecule according to claim 17 wherein said second CS protein sequence is peptide-bonded at the amino-terminus of said repeated sequence.

20.   The recombinant HBc chimer protein molecule according to claim 16 wherein said second CS protein sequence is SEQ ID NO:126 (Asn-Ala-Asn-Pro-Asn-Val-Asp-Pro).

21. The recombinant HBc chimer protein molecule according to claim 20 wherein said second CS protein sequence is peptide-bonded at the carboxy-terminus of said repeated sequence.

22. The recombinant HBc chimer protein molecule according to claim 20 wherein said second CS protein sequence is peptide-bonded at the amino-terminus of said repeated sequence.

23. The recombinant HBc chimer protein molecule according to claim 10 wherein one to three cysteine residues are present in Domain IV.

24. The recombinant HBc chimer protein molecule according to claim 23 wherein said one to three cysteine residues are present in said T cell epitope.

25. The recombinant HBc chimer protein molecule according to claim 24 wherein said T cell epitope is present and has the sequence of SEQ ID NO: 148 or 25.

26. The recombinant HBc chimer protein molecule according to claim 9 present as self-assembled particles.

27. Particles comprised of recombinant hepatitis B virus core (HBc) protein chimer molecules, said molecules having a sequence of about 155 to about 235 amino acid residues that contains four peptide-linked amino acid residue sequence

-138-

domains from the N-terminus that are denominated
Domains I, II, III and IV, wherein

(a) Domain I comprises about 71 to about 85
amino acid residues whose sequence includes at least
the sequence of the residues of position 5 through
position 75 of HBc;

(b) Domain II comprises about 18 to about
58 amino acid residues peptide-bonded to residue 75
of which (i) a sequence of HBc is present from HBc
positions 76 through 85 and (ii) a sequence of 8 to
about 48 residues that constitute a B cell epitope of
the circumsporozoite protein of a species of the
parasite *Plasmodium* that is peptide-bonded between
the HBc residues of positions 78 and 79;

(c) Domain III is an HBc sequence from
position 86 through position 135 peptide-bonded to
residue 85; and

d) Domain IV comprises (i) zero to fourteen
residues of a HBc amino acid residue sequence from
position 136 through 149 peptide-bonded to the
residue of position 135 of Domain III, (ii) zero to
three cysteine residues, (iii) fewer than three
arginine or lysine residues, or mixtures thereof
adjacent to each other, and (iv) up to 50 amino acid
residues in a sequence heterologous to HBc from
position 150 to the C-terminus, with the proviso that
at least five amino acid residues are present of the
HBc amino acid residue sequence from position 136
through 149 when (a) zero cysteine residues are
present and (b) fewer than about five heterologous
amino acid residues are present.

28. The particles according to claim 27
whose HBc chimer protein molecules have a sequence
length of about 165 to about 210 amino acid residues.

29. The particles according to claim 27
wherein said B cell epitope consists essentially of
two to about five repeats of a 4 to about 11 amino
acid residue sequence.

30. The particles according to claim 29
wherein said *Plasmodium* species is *falciparum*.

31. The particles according to claim 29
wherein said *Plasmodium* species is *vivax*.

32. The particles according to claim 27
wherein Domain I consists essentially of the HBc
sequence from position 1 through position 75.

33. The particles according to claim 27
wherein Domain II independently includes zero to
three peptide-bonded residues on either side of said
B cell epitope that are other than those of HBc or
said B cell epitope.

34. The particles according to claim 27
wherein said sequence heterologous to HBc at position
150 to the C-terminus of Domain IV comprises an amino
acid residue sequence that constitutes a T cell
epitope of the same species of *Plasmodium* as said B
cell epitope.

35. Particles comprised of recombinant
hepatitis B virus core (HBc) protein chimer

molecules, said molecules having a sequence of about
165 to about 210 amino acid residues that contains
four peptide-linked amino acid residue sequence
domains from the N-terminus that are denominated
Domains I, II, III and IV, wherein

(a) Domain I consists essentially of a
sequence of residues 1 through position 75 of HBc;

(b) Domain II comprises about 18 to about
46 amino acid residues peptide-bonded to residue 75
of which (i) 10 residues are present in a sequence of
HBc from position 76 through 85 and (ii) a sequence
of 8 to about 36 residues that constitute a B cell
epitope of the circumsporozoite (CS) protein of
Plasmodium falciparum or *Plasmodium vivax* that is
peptide-bonded between the residues of HBc positions
78 and 79, said B cell epitope being comprised of two
to about five repeats of an amino acid residue
sequence, said Domain independently including zero to
two peptide-bonded residues on either side of said B
cell epitope that are other than those of HBc or said
B cell epitope;

(c) Domain III consists essentially of the
HBc sequence from position 86 through position 135
peptide-bonded to residue 85; and

d) Domain IV comprises the HBc sequence of
residues 136 through 140 peptide-bonded to the
residue of position 135 of Domain III and (i) zero to
nine residues of a HBc amino acid residue sequence
from position 140 through 149 peptide-bonded to the
residue of position 140, (ii) zero to three cysteine
residues, (iii) fewer than three arginine or lysine
residues, or mixtures thereof adjacent to each other,
and (iv) up to 25 amino acid residues in a sequence
that constitutes a T cell epitope of the same species

of *Plasmodium* as said B cell epitope, said T cell
epitope sequence being peptide-bonded to the final
HBc amino acid residue present in a chimer molecule
or a cysteine residue.

36.   The particles according to claim 35
wherein Domain IV comprises one to a sequence of nine
amino acid residues of the HBc sequence from residue
position 141 through position 149 linked between
residue 140 of said Domain III sequence and a
Plasmodium falciparum or *Plasmodium vivax* T cell
epitope.

37.   The particles according to claim 36
wherein the nine amino acid residues of the HBc
sequence from residue position 141 through position
149 are present.

38.   The particles according to claim 35
wherein the repeated sequence of said B cell epitope
of Domain II is SEQ ID NO:152 and SEQ ID NO:153.

39.   The particles according to claim 35
wherein the repeated sequence of said B cell epitope
of Domain II is Asn-Ala-Asn-Pro, SEQ ID NO: 184.

40.   The particles according to claim 39
wherein the repeated sequence of said B cell epitope
of Domain II is repeated three or four times.

41.   The particles according to claim 40
wherein the repeated sequences are peptide-bonded to
each other without interruption.

42. The particles according to claim 41 wherein said B cell epitope includes a second CS protein sequence from the same *Plasmodium* species that is peptide-bonded to said repeated sequence.

43. The particles according to claim 42 wherein said second CS protein sequence is Asn-Val-Asp-Pro; SEQ ID NO:185.

44. The particles according to claim 43 wherein said second CS protein sequence is peptide-bonded at the carboxy-terminus of said repeated sequence.

45. The particles according to claim 43 wherein said second CS protein sequence is peptide-bonded at the amino-terminus of said repeated sequence.

46. The particles according to claim 42 wherein said second CS protein sequence is SEQ ID NO:126 (Asn-Ala-Asn-Pro-Asn-Val-Asp-Pro).

47. The particles according to claim 46 wherein said second CS protein sequence is peptide-bonded at the carboxy-terminus of said repeated sequence.

48. The particles according to claim 46 wherein said second CS protein sequence is peptide-bonded at the amino-terminus of said repeated sequence.

49. The particles according to claim 35 wherein said B cell epitope of *Plasmodium falciparum* has an amino acid residue sequence selected from the group consisting of SEQ ID NOs:1-14.

50. The particles according to claim 35 wherein said B cell epitope of *Plasmodium vivax* has an amino acid residue sequence selected from the group consisting of SEQ ID NOs:15-21.

51. The particles according to claim 35 wherein said T cell epitope of *Plasmodium falciparum* is present and has the amino acid residue sequence of SEQ ID NO:24.

52. The particles according to claim 35 wherein said T cell epitope of *Plasmodium vivax* is present and has the amino acid residue sequence of SEQ ID NO:25.

53. The particles according to claim 36 further including one to three cysteine residues in the Domain IV sequence.

54. The particles according to claim 53 having one cysteine residue in the Domain IV sequence.

55. The particles according to claim 54 wherein said one cysteine is the carboxy-terminal residue of said chimeric molecules.

56. Particles comprised of recombinant hepatitis B virus core (HBc) protein chimer

molecules, said molecules having a sequence of about
165 to about 210 amino acid residues that contain
four peptide-linked domains from the N-terminus that
are denominated Domains I, II, III and IV, wherein

(a) Domain I consists essentially of a
sequence of residues 1 through position 75 of HBc;

(b) Domain II comprises about 18 to about
46 amino acid residues peptide-bonded to residue 75
of which (i) 10 residues are present in a sequence of
HBc from position 76 through 85 and (ii) a sequence
that constitutes a B cell epitope of the
circumsporozoite (CS) protein of *Plasmodium
falciparum* or *Plasmodium vivax* is peptide-bonded
between the residues of HBc positions 78 and 79, said
B cell epitope being selected from the group
consisting of SEQ ID NOs:1-21, said Domain II
including two peptide-bonded residues on either side
of said B cell epitope that are other than those of
HBc or said B cell epitope;

(c) Domain III consists essentially of the
HBc sequence from position 86 through position 135
peptide bonded to residue 85; and

d) Domain IV comprises the sequence of HBc
residues 136-140 peptide-bonded to residue 135 plus
one to nine residues of a HBc amino acid residue
sequence from position 141 through 149 peptide-bonded
to the residue of position 140 and also peptide-
bonded to a *Plasmodium falciparum* or *Plasmodium vivax*
T cell epitope of a sequence of up to about 25 amino
acid residues that includes a cysteine residue.

57.   The particles according to claim 56
wherein Domain IV comprises nine amino acid residues
of the HBc sequence from residue position 141 through

position 149 bonded between said residue 140 and said
Plasmodium falciparum or *Plasmodium vivax* T cell
epitope.

58.   The particles according to claim 57
wherein said B cell epitope is of the CS protein of
Plasmodium falciparum that is selected from the group
consisting of SEQ ID NOs:1-14 and said *Plasmodium
falciparum* T cell epitope has the amino acid sequence
of SEQ ID NO:24.

59.   The particles according to claim 57
wherein said B cell epitope is of the CS protein of
Plasmodium vivax that is selected from the group
consisting of SEQ ID NOs:15-21 and said *Plasmodium
vivax* T cell epitope has the amino acid sequence of
SEQ ID NO:25.

60.   A vaccine or inoculum comprising an
immunogenic effective amount immunogenic particles
dissolved or dispersed in a pharmaceutically
acceptable diluent, wherein said immunogenic
particles are comprised of a plurality of recombinant
chimeric hepatitis B core (HBc) protein molecules
having a length of about 140 to about 310 amino acid
residues that contain four peptide-linked amino acid
residue sequence domains from the N-terminus that are
denominated Domains I, II, III and IV, wherein
         (a) Domain I comprises about 71 to about 85
amino acid residues whose sequence includes at least
the sequence of the residues of position 5 through
position 75 of HBc;
         (b) Domain II comprises about 18 to about
58 amino acid residues peptide-bonded to residue 75

of which (i) a sequence of HBc is present from HBc positions 76 through 85 and (ii) a sequence of 8 to about 48 residues that constitute a B cell epitope of the circumsporozoite protein of a species of the parasite *Plasmodium* that is peptide-bonded between the HBc residues of positions 78 and 79;

(c) Domain III is an HBc sequence from position 86 through position 135 peptide-bonded to residue 85; and

d) Domain IV comprises (i) zero to fourteen residues of a HBc amino acid residue sequence from position 136 through 149 peptide-bonded to the residue of position 135 of Domain III,, (ii) zero to three cysteine residues, (iii) fewer than three arginine or lysine residues, or mixtures thereof adjacent to each other, and (iv) up to 100 amino acid residues in a sequence heterologous to HBc from position 150 to the C-terminus, with the proviso that at least five amino acid residues are present of the amino acid residue sequence from position 136 through 149, when (a) zero cysteine residues are present and (b) fewer than about five heterologous amino acid residues are present.

61. The vaccine or inoculum according to claim 60 wherein said immunogenic particles are those according to claim 37.

62. The vaccine or inoculum according to claim 60 wherein said immunogenic particles are those according to claim 40.

63. The vaccine or inoculum according to claim 60 wherein said immunogenic particles are those according to claim 42.

64. The vaccine or inoculum according to claim 60 wherein said immunogenic particles are those according to claim 57.

65. The vaccine or inoculum according to claim 60 wherein said immunogenic particles are those according to claim 58.

66. The vaccine or inoculum according to claim 60 wherein said immunogenic particles are those according to claim 59.

67. The vaccine or inoculum according to claim 60 that is adapted for parenteral administration.

68. A nucleic acid that encodes a recombinant HBc protein molecule according to claim 1, or a variant, analog or complement thereof.

69. A nucleic acid that encodes a recombinant HBc protein molecule according to claim 9, or a variant, analog or complement thereof.

70. A recombinant nucleic acid molecule that comprises a vector operatively linked to a nucleic acid segment defining a gene that encodes a recombinant HBc protein molecule according to claim 1, or a varient, analog or complement thereof, and a

promoter suitable for driving the expression of the
gene in a compatible host organism.

71.   A recombinant nucleic acid molecule
that comprises a vector operatively linked to a
nucleic acid segment defining a gene that encodes a
recombinant HBc protein molecule according to claim
9, or a varient, analog or complement thereof, and a
promoter suitable for driving the expression of the
gene in a compatible host organism.

72.   A host cell transformed with a
recombinant nucleic acid molecule according to claim
70.

73.   The transformed host cell according to
claim 72 wherein said host cell is selected from the
group consisting of *E. coli*, *S. typhi*, *S. typhimurium*
and a *S. typhimurium-E. coli* hybrid.

74.   A host cell transformed with a
recombinant nucleic acid molecule according to claim
71.

75.   The transformed host cell according to
claim 74 wherein said host cell is selected from the
group consisting of *E. coli*, *S. typhi*, *S. typhimurium*
and a *S. typhimurium-E. coli* hybrid.

FIG. 1

pKK223-3   TTCACAGGAAACAGAATTCCCGGGGATCCGTCGGACCTGCAGCCAAGCTT
                                                          HindIII

                      SEQ ID NO:180

pKK223-3N   TTCACATAAGGAGGAAAAACcatggGATCCG--------AAGCTT
                                         NcoI

                      SEQ ID NO:181

FIG. 2A

Cloning Step 1

I   N   A   N   P   N   A   N   P   N   A   N   P   N   A   N   P   E   L
A ATT AAC GCT AAT CCG AAC GCT AAT CCG AAC GCT AAT CCG AAC GCT AAT CCG GAG CT
  TTG CGA TTA GGC TTG CGA TTA GGC TTG CGA TTA GGC TTG CGA TTA GGC C

| HBc 1-78 | G | C | HBc 79-149 |
| | CT TAA | TC GAG | |

HBc 1-78 + (NANP)₄
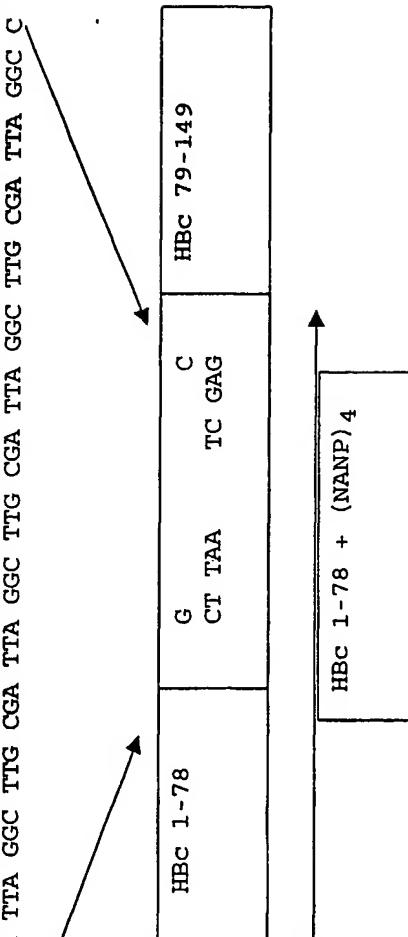
FIG. 2B

Cloning Step 2 –

```
    I   E   Y   L   N   K   I   Q   N   S   L   S   T   E   W   S   P   C   S   V   T   *
A ATT GAA TAT CTG AAC AAA ATC CAG AAC TCT CTG TCC ACC GAA TGG TCT CCG TGC TCC GTT ACC TAA AAG CT
  TAA CTT ATA GAC TTG TTT TAG GTC TTG AGA GAC AGG TGG CTT ACC AGA GGC ACG AGG CAA TGG ATT T
```

```
            G              T
         CT TAA         TC GAA
```

HBc79-149 + PF/CS326-345
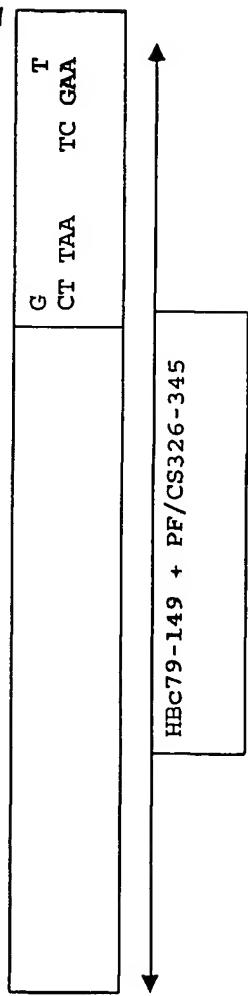
FIG. 2C

Cloning step 3
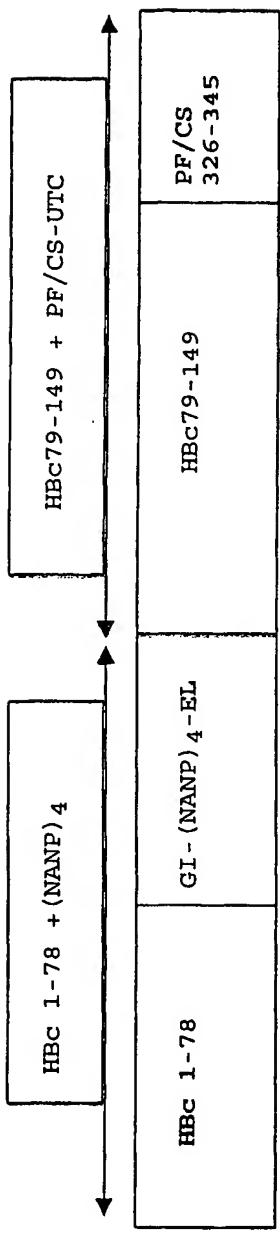
FIG. 3
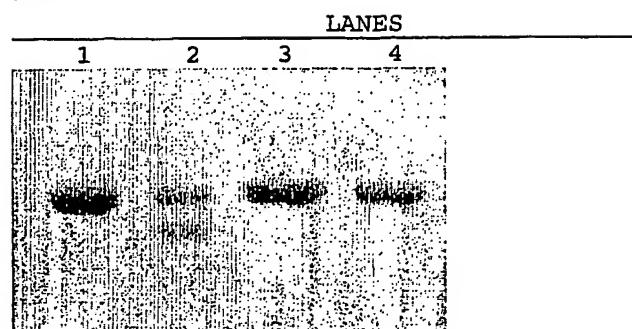
FIG. 4

FIG. 5
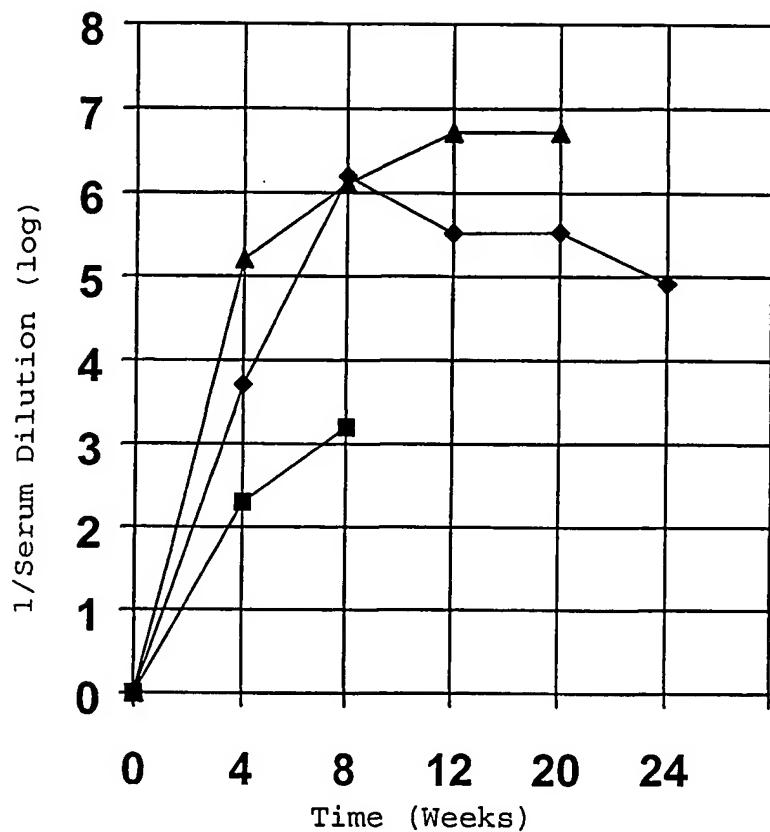
FIG. 6A

Ground Squirrel   mylfhlclvf acvpcptvqa sklclgwlwd

                   1
HBc AYW   mdidpykefg atvellsflp sdffpsvrdl ldtasalyre
HBc ADW   mdidpykefg atvellsflp sdffpsvrdl ldtasalyre
HBc ADW2   mdidpykefg atvellsflp sdffpsvrdl ldtasalyre
HBc ADYW   mdidpykefg atvellsflp sdffpsvrdl ldtaaalyrd
Woodchuck   mdidpykefg ssyqllnflp ldffpdlnal vdtatalyee
Ground Squirrel   mdidpykefg ssyqllnflp ldffpdlnal vdtaaalyee

                   41
HBc AYW   alespehcsp hhtalrqail cwgelmtlat wvgvnledpa
HBc ADW   alespehcsp hhtalrqail cwgelmtlat wvgnnlqdpa
HBc ADW2   alespehcsp hhtalrqail cwgelmtlat wvgnnledpa
HBc ADYW   alespehcsp hhtalrqail cwgdlmtlat wvgtnledpa
Woodchuck   eltgrehcsp hhtairqalv cwdeltklia wmssnitseq
Ground Squirrel   eltgrehcsp hhtairqalv cweeltrlit wmsentteev

FIG. 6B

```
    81
HBc AYW          srdlvvsyvn  tnmglkfrql  lwfhiscltf  gretvieylv
HBc ADW          srdlvvnyvn  tnmglkirql  lwfhiscltf  gretvleylv
HBc ADW2         srdlvvnyvn  tnvglkirql  lwfhiscltf  gretvleylv
HBc ADYW         srdlvvsyvn  tnvglkfrql  lwfhiscltf  gretvleylv
Woodchuck        vrtiivmhvn  dtwglkvrqs  lwfhlscltf  gqhtvqeflv
Ground Squirrel  rriivdhvnn  twglkvrqtl  wfhlscltfg  qhtvqeflvs

    121
HBc AYW          sfgvwirtpp  ayrppnapil  stlpettvvr  rrgrsprrrt
HBc ADW          sfgvwirtpp  ayrppnapil  stlpettvvr  rrdrgrsprr
HBc ADW2         sfgvwirtpp  ayrppnapil  stlpettvvr  rrdrgrsprr
HBc ADYW         sfgvwirtpp  ayrppnapil  stlpettvvr  rrgrsprrrt
Woodchuck        sfgvwirtpa  pyrppnapil  stlpehtvir  rrggarasrs
Ground Squirrel  fgvwirtpap  yrppnapils  tlpehtvirr  rggsraarsp

    161
HBc AYW          psprrrrsqs  prrrrsqsre  sqc
HBc ADW          rtpsprrrrs  qsprrrrsqs  resqc
HBc ADW2         rtpsprrrps  qsprrrrsqs  resqc
HBc ADYW         psprrrrsqs  prrrrsqsre  sqc
Woodchuck        prrrtpsprr  rrsqsprrrr  sqc
Ground Squirrel  rrrtpsprrr  rsqsprrrrs  qspasnc
```

FIG. 7

V12.Pf1			V12.Pf1(C17A)C150		
		DAYS			
0	7	14	0	7	14
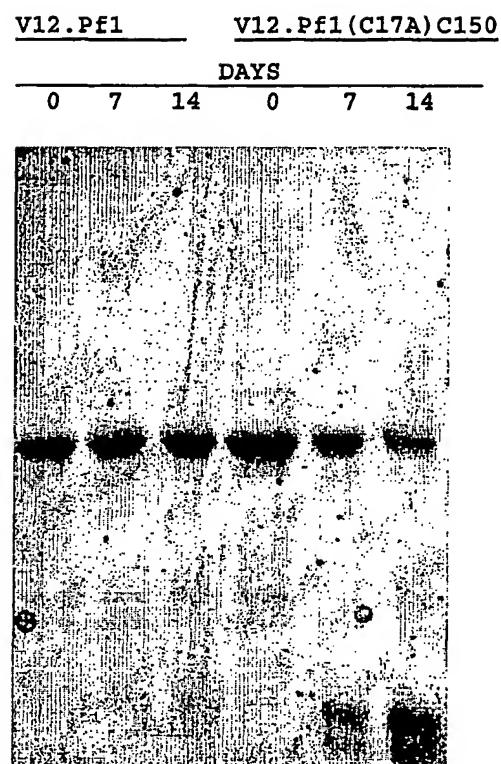
SEQUENCE LISTING

<110> Birkett, Ashley J.

<120> MALARIA IMMUNOGEN AND VACCINE

<130> 4564/83503 ICC-103.1 PCT

<140> Not Yet Assigned
<141> 2001-08-16

<150> 60/225,843
<151> 2000-08-16

<150> USSN NOT YET ASSIGND
<151> 2001-08-15

<160> 186

<170> PatentIn Ver. 2.1

<210> 1
<211> 16
<212> PRT
<213> Plasmodium falciparum

<400> 1
Asn Ala Asn Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Ala Asn Pro
 1   5   10   15

<210> 2
<211> 24
<212> PRT
<213> Plasmodium falciparum

<400> 2
Asn Ala Asn Pro Asn Val Asp Pro Asn Ala Asn Pro Asn Ala Asn Pro
 1   5   10   15

Asn Ala Asn Pro Asn Val Asp Pro
            20

<210> 3
<211> 20
<212> PRT
<213> Plasmodium falciparum

<400> 3
Asn Ala Asn Pro Asn Val Asp Pro Asn Ala Asn Pro Asn Ala Asn Pro
 1   5   10   15

Asn Ala Asn Pro
            20

<210> 4
<211> 20
<212> PRT
<213> Plasmodium falciparum

```
<400> 4
Asn Ala Asn Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Val Asp Pro
 1               5                10                15

Asn Ala Asn Pro
            20


<210> 5
<211> 28
<212> PRT
<213> Plasmodium falciparum

<400> 5
Asn Ala Asn Pro Asn Val Asp Pro Asn Ala Asn Pro Asn Ala Asn Pro
 1               5                10                15

Asn Ala Asn Pro Asn Val Asp Pro Asn Ala Asn Pro
            20                25


<210> 6
<211> 20
<212> PRT
<213> Plasmodium falciparum

<400> 6
Asn Pro Asn Val Asp Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Ala
 1               5                10                15

Asn Pro Asn Val
            20


<210> 7
<211> 22
<212> PRT
<213> Plasmodium falciparum

<400> 7
Asn Pro Asn Val Asp Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Ala
 1               5                10                15

Asn Pro Asn Val Asp Pro
            20


<210> 8
<211> 24
<212> PRT
<213> Plasmodium falciparum

<400> 8
Asn Pro Asn Val Asp Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Ala
 1               5                10                15

Asn Pro Asn Val Asp Pro Asn Ala
            20
```

```
<210> 9
<211> 18
<212> PRT
<213> Plasmodium falciparum

<400> 9
Asn Val Asp Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Ala Asn Pro
 1               5                   10                  15

Asn Val
```

```
<210> 10
<211> 20
<212> PRT
<213> Plasmodium falciparum

<400> 10
Asn Val Asp Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Ala Asn Pro
 1               5                 . 10                  15

Asn Val Asp Pro
            20
```

```
<210> 11
<211> 22
<212> PRT
<213> Plasmodium falciparum

<400> 11
Asn Val Asp Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Ala Asn Pro
 1               5                   10                  15

Asn Val Asp Pro Asn Ala
            20
```

```
<210> 12
<211> 16
<212> PRT
<213> Plasmodium falciparum

<400> 12
Asp Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Val
 1               5                   10                  15
```

```
<210> 13
<211> 18
<212> PRT
<213> Plasmodium falciparum

<400> 13
Asp Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Val
 1               5                   10                  15

Asp Pro
```

```
<210> 14
<211> 20
<212> PRT
<213> Plasmodium falciparum

<400> 14
Asp Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Val
 1               5                  10                  15

Asp Pro Asn Ala
            20


<210> 15
<211> 18
<212> PRT
<213> Plasmodium vivax

<400> 15
Asp Arg Ala Ala Gly Gln Pro Ala Gly Asp Arg Ala Asp Gly Gln Pro
 1               5                  10                  15

Ala Gly


<210> 16
<211> 36
<212> PRT
<213> Plasmodium vivax

<400> 16
Ala Asn Gly Ala Gly Asn Gln Pro Gly Ala Asn Gly Ala Gly Asp Gln
 1               5                  10                  15

Pro Gly Ala Asn Gly Ala Asp Asn Gln Pro Gly Ala Asn Gly Ala Asp
            20                  25                  30

Asp Gln Pro Gly
         35


<210> 17
<211> 18
<212> PRT
<213> Plasmodium vivax

<400> 17
Ala Asn Gly Ala Gly Asn Gln Pro Gly Ala Asn Gly Ala Gly Asp Gln
 1               5                  10                  15

Pro Gly


<210> 18
<211> 18
<212> PRT
<213> Plasmodium vivax
```

-4-

```
<400> 18
Ala Asn Gly Ala Asp Asn Gln Pro Gly Ala Asn Gly Ala Asp Asp Gln
 1               5                   10                  15

Pro Gly


<210> 19
<211> 18
<212> PRT
<213> Plasmodium vivax

<400> 19
Ala Asn Gly Ala Gly Asn Gln Pro Gly Ala Asn Gly Ala Asp Asn Gln
 1               5                   10                  15

Pro Gly


<210> 20
<211> 18
<212> PRT
<213> Plasmodium vivax

<400> 20
Ala Asn Gly Ala Asp Asn Gln Pro Gly Ala Asn Gly Ala Asp Asp Gln
 1               5                   10                  15

Pro Gly


<210> 21
<211> 22
<212> PRT
<213> Plasmodium vivax

<400> 21
Ala Pro Gly Ala Asn Gln Glu Gly Gly Ala Ala Ala Pro Gly Ala Asn
 1               5                   10                  15

Gln Glu Gly Gly Ala Ala
            20


<210> 22
<211> 16
<212> PRT
<213> Plasmodium berghei

<400> 22
Asp Pro Pro Pro Pro Asn Pro Asn Asp Pro Pro Pro Asn Pro Asn
 1               5                   10                  15


<210> 23
<211> 24
<212> PRT
<213> Plasmodium yoelii
```

<400> 23
Gln Gly Pro Gly Ala Pro Gln Gly Pro Gly Ala Pro Gln Gly Pro Gly
  1   5   10   15

Ala Pro Gln Gly Pro Gly Ala Pro
              20

<210> 24
<211> 22
<212> PRT
<213> Plasmodium falciparum

<400> 24
Gly Ile Glu Tyr Leu Asn Lys Ile Gln Asn Ser Leu Ser Thr Glu Trp
  1   5   10   15

Ser Pro Cys Ser Val Thr
              20

<210> 25
<211> 19
<212> PRT
<213> Plasmodium vivax

<400> 25
Tyr Leu Asp Lys Val Arg Ala Thr Val Gly Thr Glu Trp Thr Pro Cys
  1   5   10   15

Ser Val Thr

<210> 26
<211> 20
<212> PRT
<213> Plasmodium yoelii

<400> 26
Glu Phe Val Lys Gln Ile Ser Ser Gln Leu Thr Glu Glu Trp Ser Gln
  1   5   10   15

Cys Ser Val Thr
              20

<210> 27
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: plasmid pkk223

<400> 27
ggtgcatgca aggagatg   18

<210> 28
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:pKK223

<400> 28
gcgaagcttc ggatcccatg gttttttcct ccttatgtga aattgttatc cgctc   55

<210> 29
<211> 24
<212> DNA
<213> Hepatitis B virus

<400> 29
ttgggccatg gacatcgacc ctta   24

<210> 30
<211> 29
<212> DNA
<213> Hepatitis B virus

<400> 30
gcggaattcc ttccaaatta acacccacc   29

<210> 31
<211> 38
<212> DNA
<213> Hepatitis B virus

<400> 31
cgcgaattca aaagagctc gatccagcgt ctagagac   38

<210> 32
<211> 31
<212> DNA
<213> Hepatitis B virus

<400> 32
cgcaagctta aacaacagta gtctccggaa g   31

<210> 33
<211> 42
<212> DNA
<213> Hepatitis B virus

<400> 33
cgcaagctta gagctcttga attccaacaa cagtagtctc cg   42

<210> 34
<211> 39
<212> DNA
<213> Hepatitis B virus

```
<400> 34
cgcgaattca aaaagagctc ccagcgtcta gagacctag                    39


<210> 35
<211> 27
<212> DNA
<213> Hepatitis B virus

<400> 35
cgcgagctcc cagcgtctag agacctag                                28


<210> 36
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:pKK223-2

<400> 36
gtatcaggct gaaaatc                                            17


<210> 37
<211> 19
<212> PRT
<213> Plasmodium falciparum

<400> 37
Ile Asn Ala Asn Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Ala Asn
 1               5                   10                  15

Pro Glu Leu


<210> 38
<211> 57
<212> DNA
<213> Plasmodium falciparum

<400> 38
aattaacgct aatccgaacg ctaatccgaa cgctaatccg aacgctaatc cggagct   57


<210> 39
<211> 49
<212> DNA
<213> Plasmodium falciparum

<400> 39
ccggattagc gttcggatta gcgttcggat tagcgttcgg attagcgtt          49


<210> 40
<211> 31
<212> PRT
<213> Plasmodium falciparum
```

<400> 40
Ile Asn Ala Asn Pro Asn Val Asp Pro Asn Ala Asn Pro Asn Ala Asn
 1   5   10   15

Pro Asn Ala Asn Pro Asn Val Asp Pro Asn Ala Asn Pro Glu Leu
            20   25   30

<210> 41
<211> 93
<212> DNA
<213> Plasmodium falciparum

<400> 41
aattaacgct aatccgaacg ttgacccgaa cgctaatccg aacgctaatc cgaacgctaa 60
tccgaacgtt gacccgaacg ctaatccgga gct   93

<210> 42
<211> 91
<212> DNA
<213> Plasmodium falciparum

<400> 42
ggagctccgg attagcgttc gggtcaacgt tcggattagc gttcggatta gcgttcggat 60
tagcgttcgg gtcaacgttc ggattagcgt t   91

<210> 43
<211> 23
<212> PRT
<213> Plasmodium berghei

<400> 43
Ile Asn Ala Asn Pro Asn Val Asp Pro Asn Ala Asn Pro Asn Ala Asn
 1   5   10   15

Pro Asn Ala Asn Pro Glu Leu
            20

<210> 44
<211> 69
<212> DNA
<213> Plasmodium falciparum

<400> 44
aattaacgcg aatccgaacg tggatccgaa tgccaaccct aacgccaacc caaatgcgaa 60
cccagagct   69

<210> 45
<211> 61
<212> DNA
<213> Plasmodium falciparum

<400> 45
ctgggttcgc atttgggttg gcgttagggt tggcattcgg atccacgttc ggattcgcgt 60
t   61

```
<210> 46
<211> 23
<212> PRT
<213> Plasmodium falciparum

<400> 46
Ile Asn Ala Asn Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Val Asp
 1               5                  10                  15

Pro Asn Ala Asn Pro Glu Leu
            20


<210> 47
<211> 69
<212> DNA
<213> Plasmodium falciparum

<400> 47
aattaacgcg aatccgaatg ccaaccctaa cgccaaccca aacgtggatc cgaatgcgaa 60
cccagagct                                                        69


<210> 48
<211> 61
<212> DNA
<213> Plasmodium falciparum

<400> 48
ctgggttcgc attcggatcc acgtttgggt tggcgttagg gttggcattc ggattcgcgt 60
t                                                                61


<210> 49
<211> 31
<212> PRT
<213> Plasmodium falciparum

<400> 49
Ile Asn Ala Asn Pro Asn Val Asp Pro Asn Ala Asn Pro Asn Ala Asn
 1               5                  10                  15

Pro Asn Ala Asn Pro Asn Val Asp Pro Asn Ala Asn Pro Glu Leu
            20                  25                  30


<210> 50
<211> 93
<212> DNA
<213> Plasmodium falciparum

<400> 50
aattaacgcg aatccgaacg tggatccaaa tgccaaccct aacgctaatc caaacgccaa 60
cccgaatgtt gaccccaatg ccaatccgga gct                              93


<210> 51
<211> 85
<212> DNA
<213> Plasmodium falciparum
```

```
<400> 51
ccggattggc attggggtca acattcgggt tggcgtttgg attagcgtta gggttggcat 60
ttggatccac gttcggattc gcgtt                                       85


<210> 52
<211> 23
<212> PRT
<213> Plasmodium falciparum

<400> 52
Ile Asn Pro Asn Val Asp Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn
 1               5               10              15

Ala Asn Pro Asn Val Glu Leu
            20


<210> 53
<211> 69
<212> DNA
<213> Plasmodium falciparum

<400> 53
aattaatccg aacgtggatc caaatgccaa ccctaacgct aatccaaacg ccaacccgaa 60
tgttgagct                                                        69


<210> 54
<211> 61
<212> DNA
<213> Plasmodium falciparum

<400> 54
caacattcgg gttggcgttt ggattagcgt tagggttggc atttggatcc acgttcggat 60
t                                                                61


<210> 55
<211> 25
<212> PRT
<213> Plasmodium falciparum

<400> 55
Ile Asn Pro Asn Val Asp Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn
 1               5               10              15

Ala Asn Pro Asn Val Asp Pro Glu Leu
            20              25


<210> 56
<211> 75
<212> DNA
<213> Plasmodium falciparum

<400> 56
aattaatccg aacgtggatc caaatgccaa ccctaacgct aatccaaacg ccaacccgaa 60
tgttgaccct gagct                                                 75
```

```
<210> 57
<211> 67
<212> DNA
<213> Plasmodium falciparum

<400> 57
cagggtcaac attcgggttg gcgtttggat tagcgttagg gttggcattt ggatccacgt 60
tcggatt                                                           67


<210> 58
<211> 27
<212> PRT
<213> Plasmodium falciparum

<400> 58
Ile Asn Pro Asn Val Asp Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn
 1               5                  10                  15

Ala Asn Pro Asn Val Asp Pro Asn Ala Glu Leu
            20                  25


<210> 59
<211> 81
<212> DNA
<213> Plasmodium falciparum

<400> 59
aattaatccg aacgtggatc caaatgccaa ccctaacgct aatccaaacg ccaacccgaa 60
tgttgaccct aatgctgagc t                                           81


<210> 60
<211> 73
<212> DNA
<213> Plasmodium falciparum

<400> 60
cagcattagg gtcaacattc gggttggcgt ttggattagc gttagggttg gcatttggat 60
ccacgttcgg att                                                    73


<210> 61
<211> 21
<212> PRT
<213> Plasmodium falciparum

<400> 61
Ile Asn Val Asp Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Ala Asn
 1               5                  10                  15

Pro Asn Val Glu Leu
            20


<210> 62
<211> 63
<212> DNA
<213> Plasmodium falciparum
```

```
<400> 62
aattaacgtg gatccaaatg ccaaccctaa cgctaatcca aacgccaacc cgaatgttga  60
gct                                                                  63


<210> 63
<211> 55
<212> DNA
<213> Plasmodium falciparum

<400> 63
caacattcgg gttggcgttt ggattagcgt tagggttggc atttggatcc acgtt        55


<210> 64
<211> 23
<212> PRT
<213> Plasmodium falciparum

<400> 64
Ile Asn Val Asp Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Ala Asn
  1               5                  10                  15

Pro Asn Val Asp Pro Glu Leu
            20


<210> 65
<211> 69
<212> DNA
<213> Plasmodium falciparum

<400> 65
aattaacgtg gatccaaatg ccaaccctaa cgctaatcca aacgccaacc cgaatgttga  60
ccctgagct                                                            69


<210> 66
<211> 61
<212> DNA
<213> Plasmodium falciparum

<400> 66
cagggtcaac attcgggttg gcgtttggat tagcgttagg gttggcattt ggatccacgt  60
t                                                                    61


<210> 67
<211> 25
<212> PRT
<213> Plasmodium falciparum

<400> 67
Ile Asn Val Asp Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Ala Asn
  1               5                  10                  15

Pro Asn Val Asp Pro Asn Ala Glu Leu
            20                  25
```

```
<210> 68
<211> 75
<212> DNA
<213> Plasmodium falciparum

<400> 68
aattaacgtg gatccaaatg ccaaccctaa cgctaatcca aacgccaacc cgaatgttga 60
ccctaatgct gagct                                                  75


<210> 69
<211> 67
<212> DNA
<213> Plasmodium falciparum

<400> 69
cagcattagg gtcaacattc gggttggcgt ttggattagc gttagggttg gcatttggat 60
ccacgtt                                                           67


<210> 70
<211> 19
<212> PRT
<213> Plasmodium falciparum

<400> 70
Ile Asp Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn
  1               5                  10                  15

Val Glu Leu


<210> 71
<211> 57
<212> DNA
<213> Plasmodium falciparum

<400> 71
aattgatcca aatgccaacc ctaacgctaa tccaaacgcc aacccgaatg ttgagct    57


<210> 72
<211> 49
<212> DNA
<213> Plasmodium falciparum

<400> 72
caacattcgg gttggcgttt ggattagcgt tagggttggc atttggatc             49


<210> 73
<211> 21
<212> PRT
<213> Plasmodium falciparum

<400> 73
Ile Asp Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn
  1               5                  10                  15
```

Val Asp Pro Glu Leu
            20

            .

<210> 74
<211> 63
<212> DNA
<213> Plasmodium falciparum

<400> 74
aattgatcca aatgccaacc ctaacgctaa tccaaacgcc aacccgaatg ttgaccctga 60
gct   63

<210> 75
<211> 55
<212> DNA
<213> Plasmodium falciparum

<400> 75
cagggtcaac attcgggttg gcgtttggat tagcgttagg gttggcattt ggatc   55

<210> 76
<211> 23
<212> PRT
<213> Plasmodium falciparum

<400> 76
Ile Asp Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn Ala Asn Pro Asn
 1   5   10   15

Val Asp Pro Asn Ala Glu Leu   .
            20

<210> 77
<211> 69
<212> DNA
<213> Plasmodium falciparum

<400> 77
aattgatcca aatgccaacc ctaacgctaa tccaaacgcc aacccgaatg ttgaccctaa 60
tgccgagct   69

<210> 78   .
<211> 61
<212> DNA
<213> Plasmodium falciparum

<400> 78
cgccattagg gtcaacattc gggttggcgt ttggattagc gttagggttg gcatttggat 60
c   61

<210> 79
<211> 21
<212> PRT
<213> Plasmodium falciparum

.

<400> 79
Ile Glu Tyr Leu Asn Lys Ile Gln Asn Ser Leu Ser Thr Glu Trp Ser
  1   5   10   15

Pro Cys Ser Val Thr
              20

<210> 80
<211> 69
<212> DNA
<213> Plasmodium falciparum

<400> 80
aattgaatat ctgaacaaaa tccagaactc tctgtccacc gaatggtctc cgtgctccgt 60
tacctagta   69

<210> 81
<211> 69
<212> DNA
<213> Plasmodium falciparum

<400> 81
agcttactag gtaacggagc acggagacca ttcggtggac agagagttct ggattttgtt 60
cagatattc   69

<210> 82
<211> 24
<212> PRT
<213> Plasmodium vivax

<400> 82
Ile Pro Ala Gly Asp Arg Ala Asp Gly Gln Pro Ala Gly Asp Arg Ala
  1   5   10   15

Ala Gly Gln Pro Ala Gly Glu Leu
              20

<210> 83
<211> 72
<212> DNA
<213> Plasmodium vivax

<400> 83
aattccggct ggtgaccgtg cagatggcca gccagcgggt gaccgcgctg caggccagcc 60
ggctggcgag ct   72

<210> 84
<211> 64
<212> DNA·
<213> Plasmodium vivax

<400> 84
cgccagccgg ctggcctgca gcgcggtcac ccgctggctg gccatctgca cggtcaccag 60
ccgg   64

```
<210> 85
<211> 21
<212> PRT
<213> Plasmodium vivax

<400> 85
Ile Asp Arg Ala Ala Gly Gln Pro Ala Gly Asp Arg Ala Asp Gly Gln
 1               5                   10                  15

Pro Ala Gly Glu Leu
            20


<210> 86
<211> 63
<212> DNA
<213> Plasmodium vivax

<400> 86
aattgacaga gcagccggac aaccagcagg cgatcgagca gacggacagc ccgcagggga 60
gct                                                              63


<210> 87
<211> 55
<212> DNA
<213> Plasmodium vivax

<400> 87
cccctgcggg ctgtccgtct gctcgatcgc ctgctggttg tccggctgct ctgtc      55


<210> 88
<211> 21
<212> PRT
<213> Plasmodium falciparum

<400> 88
Ile Ala Asn Gly Ala Gly Asn Gln Pro Gly Ala Asn Gly Ala Gly Asp
 1               5                   10                  15

Gln Pro Gly Glu Leu
            20


<210> 89
<211> 63
<212> DNA
<213> Plasmodium vivax

<400> 89
aattgcgaac ggcgccggta atcagccggg ggcaaacggc gcgggtgatc aaccagggga 60
gct                                                              63


<210> 90
<211> 55
<212> DNA
<213> Plasmodium vivax
```

<400> 90
cccctggttg atcacccgcg ccgtttgccc ccggctgatt accggcgccg ttcgc   55

<210> 91
<211> 21
<212> PRT
<213> Plasmodium vivax

<400> 91
Ile Ala Asn Gly Ala Asp Asn Gln Pro Gly Ala Asn Gly Ala Asp Asp
 1   5   10   15

Gln Pro Gly Glu Leu
            20

<210> 92
<211> 63
<212> DNA
<213> Plasmodium vivax

<400> 92
aattgcgaac ggcgccgata atcagccggg tgcaaacggg gcggatgacc aaccaggcga 60
gct   63

<210> 93
<211> 55
<212> DNA
<213> Plasmodium vivax

<400> 93
cgcctggttg gtcatccgcc ccgtttgcac ccggctgatt atcggcgccg ttcgc   55

<210> 94
<211> 39
<212> PRT
<213> Plasmodium vivax

<400> 94
Ile Ala Asn Gly Ala Gly Asn Gln Pro Gly Ala Asn Gly Ala Gly Asp
 1   5   10   15

Gln Pro Gly Ala Asn Gly Ala Asp Asn Gln Pro Gly Ala Asn Gly Ala
            20   25   30

Asp Asp Gln Pro Gly Glu Leu
            35

<210> 95
<211> 117
<212> DNA
<213> Plasmodium vivax

<400> 95
aattgcgaac ggcgccggta atcagccggg agcaaacggc gcggggggatc aaccaggcgc 60
caatggtgca gacaaccagc ctggggcgaa tggagccgat gaccaacccg gcgagct   117

<210> 96
<211> 109
<212> DNA
<213> Plasmodium vivax

<400> 96
cgccgggttg gtcatcggct ccattcgccc caggctggtt gtctgcacca ttggcgcctg 60
gttgatcccc cgcgccgttt gctcccggct gattaccggc gccgttcgc   109

<210> 97
<211> 25
<212> PRT
<213> Plasmodium vivax

<400> 97
Ile Ala Pro Gly Ala Asn Gln Glu Gly Gly Ala Ala Ala Pro Gly Ala
 1   5   10   15

Asn Gln Glu Gly Gly Ala Ala Glu Leu
            20   25

<210> 98
<211> 75
<212> DNA
<213> Plasmodium vivax

<400> 98
aattgcgccg ggcgccaacc aggaaggtgg ggctgcagcg ccaggagcca atcaagaagg 60
cggtgcagcg gagct   75

<210> 99
<211> 67
<212> DNA
<213> Plasmodium vivax

<400> 99
ccgctgcacc gccttcttga ttggctcctg gcgctgcagc cccaccttcc tggttggcgc 60
ccggcgc   67

<210> 100
<211> 21
<212> PRT
<213> Plasmodium vivax

<400> 100
Ile Glu Tyr Leu Asp Lys Val Arg Ala Thr Val Gly Thr Glu Trp Thr
 1   5   10   15

Pro Cys Ser Val Thr
            20

<210> 101
<211> 69
<212> DNA
<213> Plasmodium vinckei

<400> 101
aattgaatat ctggataaag tgcgtgcgac cgttggcacg gaatggactc cgtgcagcgt 60
gacctaata   69

<210> 102
<211> 69
<212> DNA
<213> Plasmodium vivax

<400> 102
agcttattag gtcacgctcg acggagtcca ttccgtgcca acggtcgcac gcactttatc 60
cagatattc   69

<210> 103
<211> 10
<212> PRT
<213> Plasmodium falciparum

<400> 103
Thr Val Ser Ala Pro Ser Trp Glu Thr Ser
 1   5   10

<210> 104
<211> 42
<212> DNA
<213> Plasmodium falciparum

<400> 104
gccaagctta ctaggtaacg gaggccggag accattcggt gg   42

<210> 105
<211> 6
<212> PRT
<213> Hepatitis B virus

<400> 105
Met Asp Ile Asp Pro Tyr
 1   5

<210> 106
<211> 8
<212> PRT
<213> Hepatitis B virus

<400> 106
Cys Val Val Thr Thr Glu Pro Leu
 1   5

<210> 107
<211> 37
<212> DNA
<213> Hepatitis B virus

<400> 107
cgcaagctta ctagcaaaca acagtagtct ccggaag   37

<210> 108
<211> 7
<212> PRT
<213> Hepatitis B virus

<400> 108
Pro Leu Thr Ser Leu Ile Pro
  1   5

<210> 109
<211> 32
<212> DNA
<213> Hepatitis B virus

<400> 109
cgcaagctta cggaagtgtt gataggatag gg   32

<210> 110
<211> 8
<212> PRT
<213> Hepatitis B virus

<400> 110
Thr Ser Leu Ile Pro Ala Asn Pro
  1   5

<210> 111
<211> 34
<212> DNA
<213> Hepatitis B virus

<400> 111
cgcaagctta tgttgatagg atagggcat ttgg   34

<210> 112
<211> 7
<212> PRT
<213> Hepatica americana

<400> 112
Leu Ile Pro Ala Asn Pro Pro
  1   5

<210> 113
<211> 31
<212> DNA
<213> Hepatitis B virus

<400> 113
cgcaagctta taggataggg gcatttggtg g   31

```
<210> 114
<211> 6
<212> PRT
<213> Hepatitis B virus

<400> 114
Ile Pro Ala Asn Pro Pro
  1                 5


<210> 115
<211> 28
<212> DNA
<213> Hepatitis B virus

<400> 115
gcgaagctta datagggca tttggtgg                            28


<210> 116
<211> 6
<212> PRT
<213> Hepatitis B virus

<400> 116
Pro Ala Asn Pro Pro Arg
  1                 5


<210> 117
<211> 28
<212> DNA
<213> Hepatitis B virus

<400> 117
cgcaagctta aggggcattt ggtggtct                           28


<210> 118
<211> 7
<212> PRT
<213> Hepatitis B virus

<400> 118
Cys Pro Ala Asn Pro Pro Arg
  1                 5


<210> 119
<211> 31
<212> DNA
<213> Hepatitis B virus

<400> 119
gcgaagctta gcaaggggca tttggtggtc t                       31


<210> 120
<211> 7
<212> PRT
<213> Hepatitis B virus
```

```
<400> 120
Ala Asn Pro Pro Arg Tyr Ala
  1               5


<210> 121
<211> 30
<212> DNA
<213> Hepatitis B virus

<400> 121
gcgaagctta ggcatttggt ggtctatagc                                      30


<210> 122
<211> 8
<212> PRT
<213> Hepatitis B virus

<400> 122
Cys Ala Asn Pro Pro Arg Tyr Ala
  1               5


<210> 123
<211> 32
<212> DNA
<213> Hepatitis B virus

<400> 123
gcgaagctta gcaggcattt ggtggtctat aa                                   32


<210> 124
<211> 7
<212> PRT
<213> Hepatitis B virus

<400> 124
Asn Pro Pro Arg Tyr Ala Pro
  1               5


<210> 125
<211> 31
<212> DNA
<213> Hepatitis B virus

<400> 125
cgcaagctta atttggtggt ctataagctg g                                    31


<210> 126
<211> 8
<212> PRT
<213> Plasmodium falciparum

<400> 126
Asn Ala Asn Pro Asn Val Asp Pro
  1               5
```

```
<210> 127
<211> 6
<212> PRT
<213> Homo sapiens

<400> 127
Asn Tyr Lys Lys Pro Lys
  1               5


<210> 128
<211> 7
<212> PRT
<213> Homo sapiens

<400> 128
Lys Arg Gly Pro Arg Thr His
  1               5


<210> 129
<211> 21
<212> PRT
<213> Homo sapiens

<400> 129
Leu His Pro Asp Glu Thr Lys Asn Met Leu Glu Met Ile Phe Thr Pro
  1               5                  10                  15

Arg Asn Ser Asp Arg
                20


<210> 130
<211> 5
<212> PRT
<213> Human immunodeficiency virus type 1

<400> 130
Arg Ile Lys Gln Ile
  1               5


<210> 131
<211> 11
<212> PRT
<213> Human immunodeficiency virus type 1

<400> 131
Arg Ile Lys Gln Ile Gly Met Pro Gly Gly Lys
  1               5                  10


<210> 132
<211> 10
<212> PRT
<213> Human immunodeficiency virus type 1
```

<400> 132
Leu Leu Glu Leu Asp Lys Trp Ala Ser Leu
 1   5   10

<210> 133
<211> 14
<212> PRT
<213> Human immunodeficiency virus type 1

<400> 133
Glu Gln Glu Leu Leu Glu Leu Asp Lys Trp Ala Ser Leu Trp
 1   5   10

<210> 134
<211> 33
<212> PRT
<213> Human immunodeficiency virus type 1

<400> 134
Val Gln Gln Gln Asn Asn Leu Leu Arg Ala Ile Glu Ala Gln Gln His
 1   5   10   15

Leu Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Ala Arg Ile
            20   25   30

Leu

<210> 135
<211> 16
<212> PRT
<213> Human immunodeficiency virus type 1

<400> 135
His Leu Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Ala Arg
 1   5   10   15

<210> 136
<211> 36
<212> PRT
<213> Human immunodeficiency virus

<400> 136
Tyr Thr His Ile Ile Tyr Ser Leu Ile Glu Gln Ser Gln Asn Gln Gln
 1   5   10   15

Glu Lys Asn Glu Gln Glu Leu Leu Ala Leu Asp Lys Trp Ala Ser Leu
            20   25   30

Trp Asn Trp Phe
            35

<210> 137
<211> 26
<212> PRT
<213> Human immunodeficiency virus type 1

<400> 137
Tyr Thr His Ile Ile Tyr Ser Leu Ile Glu Gln Ser Gln Asn Gln Gln
 1   5   10   15

Glu Lys Asn Glu Gln Glu Leu Leu Glu Leu
            20   25

<210> 138
<211> 19
<212> PRT
<213> Homo sapiens

<400> 138
Gly Arg Glu Arg Arg Pro Arg Leu Ser Asp Arg Pro Gln Leu Pro Tyr
 1   5   10   15

Leu Glu Ala

<210> 139
<211> 20
<212> PRT
<213> Homo sapiens

<400> 139
Arg Glu Gln Arg Arg Phe Ser Val Ser Thr Leu Arg Asn Leu Gly Leu
 1   5   10   15

Gly Lys Lys Ser
            20

<210> 140
<211> 18
<212> PRT
<213> Plasmodium yoelii

<400> 140
Pro Asn Lys Leu Pro Arg Ser Thr Ala Val Val His Gln Leu Lys Arg
 1   5   10   15

Lys His

<210> 141
<211> 11
<212> PRT
<213> Plasmodium yoelii

<400> 141
Thr Ala Val Val His Gln Leu Lys Arg Lys His .
 1   5   10

```
<210> 142
<211> 22
<212> PRT
<213> Plasmodium vivax

<400> 142
Pro Ala Gly Asp Arg Ala Asp Gly Gln Pro Ala Gly Asp Arg Ala Ala
 1               5                  10                  15

Ala Gly Gln Pro Ala Gly
            20


<210> 143
<211> 12
<212> PRT
<213> Avian leukosis virus

<400> 143
Asn Gln Ser Trp Thr Met Val Ser Pro Ile Asn Val
 1               5                  10


<210> 144
<211> 16
<212> PRT
<213> Avian leukosis virus

<400> 144
Met Ile Lys Asn Gly Thr Lys Arg Thr Ala Val Thr Phe Gly Ser Val
 1               5                  10                  15


<210> 145
<211> 19
<212> PRT
<213> Foot-and-mouth disease virus

<400> 145
Pro Asn Leu Arg Gly Asp Leu Gln Val Leu Ala Gln Lys Val Ala Arg
 1               5                  10                  15

Thr Leu Pro


<210> 146
<211> 26
<212> PRT
<213> Foot-and-mouth disease virus

<400> 146
Arg Tyr Asn Arg Asn Ala Val Pro Asn Leu Arg Gly Asp Leu Gln Val
 1               5                  10                  15

Leu Ala Gln Lys Val Ala Arg Thr Leu Pro
            20                  25
```

```
<210> 147
<211> 34
<212> PRT
<213> Hepatitis B virus

<400> 147
Arg Arg Arg Gly Arg Ser Pro Arg Arg Arg Thr Pro Ser Pro Arg Arg
  1               5                  10                  15

Arg Arg Ser Gln Ser Pro Arg Arg Arg Ser Gln Ser Arg Glu Ser
            20                  25                  30

Gln Cys


<210> 148
<211> 20
<212> PRT
<213> Plasmodium falciparum       :

<400> 148
Glu Tyr Leu Asn Lys Ile Gln Asn Ser Leu Ser Thr Glu Trp Ser Pro
  1               5                  10                  15

Cys Ser Val Thr
            20


<210> 149
<211> 20
<212> PRT
<213> Plasmodium falciparum

<400> 149
Glu Tyr Leu Asn Lys Ile Gln Asn Ser Leu Ser Thr Glu Trp Ser Pro
  1               5                  10                  15

Ala Ser Val Thr
            20


<210> 150
<211> 18
<212> PRT
<213> Plasmodium vivax

<400> 150
Asp Arg Ala Ala Gly Gln Pro Ala Gly Asp Arg Ala Asp Gly Gln Pro
  1               5                  10                  15

Ala Gly


<210> 151
<211> 36
<212> PRT
<213> Plasmodium vivax
```

<400> 151
Ala Asn Gly Ala Gly Asn Gln Pro Gly Ala Asn Gly Ala Gly Asp Gln
 1   5   10   15

Pro Gly Ala Asn Gly Ala Asp Asn Gln Pro Gly Ala Asn Gly Ala Asp
            20   25   30

Asp Gln Pro Gly
        35

<210> 152
<211> 9
<212> PRT
<213> Plasmodium vivax

<400> 152
Asp Arg Ala Ala Gly Gln Pro Ala Gly
 1   5

<210> 153
<211> 9
<212> PRT
<213> Plasmodium vivax

<400> 153
Asp Arg Ala Asp Gly Gln Pro Ala Gly
 1   5

<210> 154
<211> 9
<212> PRT
<213> Plasmodium vivax

<400> 154
Ala Asn Gly Ala Gly Asn Gln Pro Gly
 1   5

<210> 155
<211> 9
<212> PRT
<213> Plasmodium vivax

<400> 155
Ala Asn Gly Ala Gly Asp Gln Pro Gly
 1   5

<210> 156
<211> 9
<212> PRT
<213> Plasmodium vivax

<400> 156
Ala Asn Gly Ala Asp Asn Gln Pro Gly
 1   5

<210> 157
<211> 9
<212> PRT
<213> Plasmodium vivax

<400> 157
Ala Asn Gly Ala Asp Asp Gln Pro Gly
  1   5

<210> 158
<211> 11
<212> PRT
<213> Plasmodium vivax

<400> 158
Ala Pro Gly Ala Asn Gln Glu Gly Gly Ala Ala
  1   5   10

<210> 159
<211> 21
<212> PRT
<213> Plasmodium vivax

<400> 159
Pro Ala Gly Asp Arg Ala Asp Gly Gln Pro Ala Gly Asp Arg Ala Ala
  1   5   10   15

Gly Gln Pro Ala Gly
            20

<210> 160
<211> 18
<212> PRT
<213> Plasmodium vivax

<400> 160
Ala Asn Gly Ala Gly Asn Gln Pro Gly Ala Asn Gly Ala Gly Asp Gln
  1   5   10   15

Pro Gly

<210> 161
<211> 19
<212> PRT
<213> Plasmodium vivax

<400> 161
Gln Ala Asn Gly Ala Asp Asn Gln Pro Gly Ala Asn Gly Ala Asp Asp
  1   5   10   15

Gln Pro Gly

```
<210> 162
<211> 44
<212> DNA
<213> Plasmodium vivax

<400> 162
cgcgaattca agcgaacggc gccgataatc agccggcggg tgca                    44


<210> 163
<211> 22
<212> PRT
<213> Plasmodium vivax

<400> 163
Ala Pro Gly Ala Asn Gln Glu Gly Gly Ala Ala Ala Pro Gly Ala Asn
 1               5                  10                  15

Gln Glu Gly Gly Ala Ala
            20


<210> 164
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: modified
      portion of Hepatitis B core

<400> 164
Cys Val Val Thr Thr Glu Pro
 1               5


<210> 165
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: modified
      portion of Hepatitis B core

<400> 165
gcaagcttac tattgaattc cgcaaacaac agtagtctcc gg                      42


<210> 166
<211> 26
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: modified
      portion of Hepatitis B core

<400> 166
Thr Thr Val Val Gly Ile Glu Tyr Leu Asn Lys Ile Gln Asn Ser Leu
 1               5                  10                  15
```

Ser Thr Glu Trp Ser Pro Cys Ser Val Thr
            20   25

<210> 167
<211> 27
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: modified
      portion of Hepatitis B core

<400> 167
Thr Thr Val Val Cys Gly Ile Glu Tyr Leu Asn Lys Ile Gln Asn Ser
  1   5   10   15

Leu Ser Thr Glu Trp Ser Pro Ala Ser Val Thr
            20   25

<210> 168
<211> 217
<212> PRT
<213> Spermophilus variegatus

<400> 168
Met Tyr Leu Phe His Leu Cys Leu Val Phe Ala Cys Val Pro Cys Pro
  1   5   10   15

Thr Val Gln Ala Ser Lys Leu Cys Leu Gly Trp Leu Trp Asp Met Asp
            20   25   30

Ile Asp Pro Tyr Lys Glu Phe Gly Ser Ser Tyr Gln Leu Leu Asn Phe
            35   40   45

Leu Pro Leu Asp Phe Phe Pro Asp Leu Asn Ala Leu Val Asp Thr Ala
        50   55   60

Ala Ala Leu Tyr Glu Glu Glu Leu Thr Gly Arg Glu His Cys Ser Pro
 65   70   75   80

His His Thr Ala Ile Arg Gln Ala Leu Val Cys Trp Glu Glu Leu Thr
                85   90   95

Arg Leu Ile Thr Trp Met Ser Glu Asn Thr Thr Glu Glu Val Arg Arg
            100   105   110

Ile Ile Val Asp His Val Asn Asn Thr Trp Gly Leu Lys Val Arg Gln
            115   120   125

Thr Leu Trp Phe His Leu Ser Cys Leu Thr Phe Gly Gly His Thr Val
        130   135   140

Gln Glu Phe Leu Val Ser Phe Gly Val Trp Ile Arg Thr Pro Ala Pro
145   150   155   160

Tyr Arg Pro Pro Asn Ala Pro Ile Leu Ser Thr Leu Pro Glu His Thr
                165   170   175

Val Ile Arg Arg Arg Gly Gly Ser Arg Ala Ala Arg Ser Pro Arg Arg
              180   185   190

Arg Thr Pro Ser Pro Arg Arg Arg Ser Gln Ser Pro Arg Arg Arg
          195   200   205

Arg Ser Gln Ser Pro Ala Ser Asn Cys
    210   215

<210> 169
<211> 651
<212> DNA
<213> Spermophilus variegatus

<400> 169
atgtatcttt ttcacctgtg ccttgttttt gcctgtgttc catgtcctac tgttcaagcc 60
tccaagctgt gccttggatg gctttgggac atggacatag atccctataa agaatttggt 120
tcttcttatc agttgttgaa ttttcttcct ttggactttt ttcctgatct caatgcattg 180
gtggacactg ctgctgctct ttatgaagaa gaattaacag gtagggagca ttgttctcct 240
catcatactg ctattagaca ggccttagtg tgttgggaag aattaactag attaattaca 300
tggatgagtg aaaatacaac agaagaagtt agaagaatta ttgttgatca tgtcaataat 360
acttggggac ttaaagtaag acagacttta tggtttcatt tatcatgtct tacttttgga 420
caacacacag ttcaagaatt tttggttagt tttggagtat ggattagaac tccagctcct 480
tatagaccac ctaatgcacc cattttatca actcttccgg aacatacagt cattaggaga 540
agaggaggtt caagagctgc taggtccccc cgaagacgca ctccctctcc tcgcaggaga 600
aggtctcaat caccgcgtcg cagacgctct caatctccag cttccaactg c 651

<210> 170
<211> 183
<212> PRT
<213> Hepatitis B virus

<400> 170
Met Asp Ile Asp Pro Tyr Lys Glu Phe Gly Ala Thr Val Glu Leu Leu
  1   5   10   15

Ser Phe Leu Pro Ser Asp Phe Phe Pro Ser Val Arg Asp Leu Leu Asp
              20   25   30

Thr Ala Ser Ala Leu Tyr Arg Glu Ala Leu Glu Ser Pro Glu His Cys
          35   40   45

Ser Pro His His Thr Ala Leu Arg Gln Ala Ile Leu Cys Trp Gly Glu
      50   55   60

Leu Met Thr Leu Ala Thr Trp Val Gly Val Asn Leu Glu Asp Pro Ala
65   70   75   80

Ser Arg Asp Leu Val Val Ser Tyr Val Asn Thr Asn Met Gly Leu Lys
                  85   90   95

Phe Arg Gln Leu Leu Trp Phe His Ile Ser Cys Leu Thr Phe Gly Arg
              100   105   110

Glu Thr Val Ile Glu Tyr Leu Val Ser Phe Gly Val Trp Ile Arg Thr
          115   120   125

Pro Pro Ala Tyr Arg Pro Pro Asn Ala Pro Ile Leu Ser Thr Leu Pro
      130   135   140

```
Glu Thr Thr Val Val Arg Arg Arg Gly Arg Ser Pro Arg Arg Arg Thr
145             150             155             160

Pro Ser Pro Arg Arg Arg Arg Ser Gln Ser Pro Arg Arg Arg Arg Ser
            165             170             175

Gln Ser Arg Glu Ser Gln Cys
        180
```

```
<210> 171
<211> 185
<212> PRT
<213> Hepatitis B virus

<400> 171
Met Asp Ile Asp Pro Tyr Lys Glu Phe Gly Ala Thr Val Glu Leu Leu
 1           5               10              15

Ser Phe Leu Pro Ser Asp Phe Phe Pro Ser Val Arg Asp Leu Leu Asp
            20              25              30

Thr Ala Ser Ala Leu Tyr Arg Glu Ala Leu Glu Ser Pro Glu His Cys
        35              40              45

Ser Pro His His Thr Ala Leu Arg Gln Ala Ile Leu Cys Trp Gly Glu
     50              55              60

Leu Met Thr Leu Ala Thr Trp Val Gly Asn Asn Leu Gln Asp Pro Ala
 65              70              75              80

Ser Arg Asp Leu Val Val Asn Tyr Val Asn Thr Asn Met Gly Leu Lys
            85              90              95

Ile Arg Gln Leu Leu Trp Phe His Ile Ser Cys Leu Thr Phe Gly Arg
            100             105             110

Glu Thr Val Leu Glu Tyr Leu Val Ser Phe Gly Val Trp Ile Arg Thr
            115             120             125

Pro Pro Ala Tyr Arg Pro Pro Asn Ala Pro Ile Leu Ser Thr Leu Pro
    130             135             140

Glu Thr Thr Val Val Arg Arg Arg Asp Arg Gly Arg Ser Pro Arg Arg
145             150             155             160

Arg Thr Pro Ser Pro Arg Arg Arg Arg Ser Gln Ser Pro Arg Arg Arg
            165             170             175

Arg Ser Gln Ser Arg Glu Ser Gln Cys
            180             185
```

```
<210> 172
<211> 185
<212> PRT
<213> Hepatitis B virus
```

<400> 172
Met Asp Ile Asp Pro Tyr Lys Glu Phe Gly Ala Thr Val Glu Leu Leu
 1   5   10   15

Ser Phe Leu Pro Ser Asp Phe Phe Pro Ser Val Arg Asp Leu Leu Asp
                20   25   30

Thr Ala Ser Ala Leu Tyr Arg Glu Ala Leu Glu Ser Pro Glu His Cys
            35   40   45

Ser Pro His His Thr Ala Leu Arg Gln Ala Ile Leu Cys Trp Gly Glu
        50   55   60

Leu Met Thr Leu Ala Thr Trp Val Gly Asn Asn Leu Glu Asp Pro Ala
65   70   75   80

Ser Arg Asp Leu Val Val Asn Tyr Val Asn Thr Asn Val Gly Leu Lys
                85   90   95

Ile Arg Gln Leu Leu Trp Phe His Ile Ser Cys Leu Thr Phe Gly Arg
            100   105   110

Glu Thr Val Leu Glu Tyr Leu Val Ser Phe Gly Val Trp Ile Arg Thr
            115   120   125

Pro Pro Ala Tyr Arg Pro Pro Asn Ala Pro Ile Leu Ser Thr Leu Pro
    130   135   140

Glu Thr Thr Val Val Arg Arg Arg Asp Arg Gly Arg Ser Pro Arg Arg
145   150   155   160

Arg Thr Pro Ser Pro Arg Arg Arg Pro Ser Gln Ser Pro Arg Arg Arg
                165   170   175

Arg Ser Gln Ser Arg Glu Ser Gln Cys
            180   185

<210> 173
<211> 183
<212> PRT
<213> Hepatitis B virus

<400> 173
Met Asp Ile Asp Pro Tyr Lys Glu Phe Gly Ala Thr Val Glu Leu Leu
 1   5   10   15

Ser Phe Leu Pro Ser Asp Phe Phe Pro Ser Val Arg Asp Leu Leu Asp
                20   25   30

Thr Ala Ala Ala Leu Tyr Arg Asp Ala Leu Glu Ser Pro Glu His Cys
            35   40   45

Ser Pro His His Thr Ala Leu Arg Gln Ala Ile Leu Cys Trp Gly Asp
        50   55   60

Leu Met Thr Leu Ala Thr Trp Val Gly Thr Asn Leu Glu Asp Pro Ala
65   70   75   80

Ser Arg Asp Leu Val Val Ser Tyr Val Asn Thr Asn Val Gly Leu Lys
                85   90   95

Phe Arg Gln Leu Leu Trp Phe His Ile Ser Cys Leu Thr Phe Gly Arg
            100   105   110

Glu Thr Val Leu Glu Tyr Leu Val Ser Phe Gly Val Trp Ile Arg Thr
            115   120   125

Pro Pro Ala Tyr Arg Pro Pro Asn Ala Pro Ile Leu Ser Thr Leu Pro
    130   135   140

Glu Thr Thr Val Val Arg Arg Arg Gly Arg Ser Pro Arg Arg Arg Thr
145   150   155   160

Pro Ser Pro Arg Arg Arg Arg Ser Gln Ser Pro Arg Arg Arg Arg Ser
            165   170   175

Gln Ser Arg Glu Ser Gln Cys
            180

<210> 174
<211> 183
<212> PRT
<213> Marmota monax

<400> 174
Met Asp Ile Asp Pro Tyr Lys Glu Phe Gly Ser Ser Tyr Gln Leu Leu
  1   5   10   15

Asn Phe Leu Pro Leu Asp Phe Phe Pro Asp Leu Asn Ala Leu Val Asp
            20   25   30

Thr Ala Thr Ala Leu Tyr Glu Glu Glu Leu Thr Gly Arg Glu His Cys
        35   40   45

Ser Pro His His Thr Ala Ile Arg Gln Ala Leu Val Cys Trp Asp Glu
    50   55   60

Leu Thr Lys Leu Ile Ala Trp Met Ser Ser Asn Ile Thr Ser Glu Gln
65   70   75   80

Val Arg Thr Ile Ile Val Asn His Val Asn Asp Thr Trp Gly Leu Lys
            85   90   95

Val Arg Gln Ser Leu Trp Phe His Leu Ser Cys Leu Thr Phe Gly Gln
            100   105   110

His Thr Val Gln Glu Phe Leu Val Ser Phe Gly Val Trp Ile Arg Thr
    115   120   125

Pro Ala Pro Tyr Arg Pro Pro Asn Ala Pro Ile Leu Ser Thr Leu Pro
    130   135   140

Glu His Thr Val Ile Arg Arg Arg Gly Gly Ala Arg Ala Ser Arg Ser
145   150   155   160

Pro Arg Arg Arg Thr Pro Ser Pro Arg Arg Arg Ser Gln Ser Pro
            165   170   175

Arg Arg Arg Arg Ser Gln Cys
            180

<210> 175
<211> 549
<212> DNA
<213> Hepatitis B virus

<400> 175
atggacatcg acccttataa agaatttgga gctactgtgg agttactctc gtttttgcct 60
tctgacttct ttccttcagt acgagatctt ctagataccg cctcagctct gtatcgggaa 120
gccttagagt ctcctgagca ttgttcacct caccatactg cactcaggca agcaattctt 180
tgctgggggg aactaatgac tctagctacc tgggtgggtg ttaatttgga agatccagcg 240
tctagagacc tagtagtcag ttatgtcaac actaatatgg gcctaaagtt caggcaactc 300
ttgtggtttc acatttcttg tctcactttt ggaagagaaa cagttataga gtatttggtg 360
tctttcggag tgtggattcg cactcctcca gcttatagac caccaaatgc ccctatccta 420
tcaacacttc cggagactac tgttgttaga cgacgaggca ggtcccctag aagaagaact 480
ccctcgcctc gcagacgaag gtctcaatcg ccgcgtcgca gaagatctca atctcgggaa 540
tctcaatgt 549

<210> 176
<211> 555
<212> DNA
<213> Hepatitis B virus

<400> 176
atggacattg acccttataa agaatttgga gctactgtgg agttactctc gtttttgcct 60
tctgacttct ttccttccgt acgagatctc ctagacaccg cctcagctct gtatcgagaa 120
gccttagagt ctcctgagca ttgctcacct caccatactg cactcaggca agccattctc 180
tgctgggggg aattgatgac tctagctacc tgggtgggta ataatttgca agatccagca 240
tccagagatc tagtagtcaa ttatgttaat actaacatgg gtttaaagat caggcaacta 300
ttgtggtttc atatatcttg cctactttt ggaagagaga ctgtacttga atatttggtc 360
tctttcggag tgtggattcg cactcctcca gcctatagac caccaaatgc ccctatctta 420
tcaacacttc cggaaactac tgttgttaga cgacgggacc gaggcaggtc ccctagaaga 480
agaactccct cgcctcgcag acgcagatct caatcgccgc gtcgcagaag atctcaatct 540
cgggaatctc aatgt 555

<210> 177
<211> 555
<212> DNA
<213> Hepatitis B virus

<400> 177
atggacattg acccttataa agaatttgga gctactgtgg agttactctc gtttttgcct 60
tctgacttct ttccttccgt cagagatctc ctagacaccg cctcagctct gtatcgagaa 120
gccttagagt ctcctgagca ttgctcacct caccatactg cactcaggca agccattctc 180
tgctgggggg aattgatgac tctagctacc tgggtgggta ataatttgga agatccagca 240
tctagggatc ttgtagtaaa ttatgttaat actaacatgg gtttaaagat caggcaacta 300
ttgtggtttc atatatcttg cctactttt ggaagagaga ctgtacttga atatttggtc 360
tctttcggag tgtggattcg cactcctcca gcctatagac caccaaatgc ccctatctta 420
tcaacacttc cggaaactac tgttgttaga cgacgggacc gaggcaggtc ccctagaaga 480
agaactccct cgcctcgcag acgcagatct ccatcgccgc gtcgcagaag atctcaatct 540
cgggaatctc aatgt 555

<210> 178
<211> 549
<212> DNA
<213> Hepatitis B virus

```
<400> 178
atggacattg acccttataa agaatttgga gctactgtgg agttactctc gttttgcct 60
tctgacttct ttccttccgt acgagatctt ctagataccg ccgcagctct gtatcgggat 120
gccttagagt ctcctgagca ttgttcacct caccatactg cactcaggca agcaattctt 180
tgctggggag acttaatgac tctagctacc tgggtgggta ctaatttaga agatccagca 240
tctagggacc tagtagtcag ttatgtcaac actaatgtgg gcctaaagtt cagacaatta 300
ttgtggtttc acatttcttg tctcactttt ggaagagaaa cggttctaga gtatttggtg 360
tcttttggag tgtggattcg cactcctcca gcttatagac caccaaatgc ccctatccta 420
tcaacgcttc cggagactac tgttgttaga cgacgaggca ggtcccctag aagaagaact 480
ccctcgcctc gcagacgaag atctcaatcg ccgcgtcgca gaagatctca atctcgggaa 540
tctcaatgt                                                       549
```

```
<210> 179
<211> 549
<212> DNA
<213> Marmota monax
```

```
<400> 179
atggacattg acccttataa agaatttgga gctactgtgg agttactctc gttttgcct 60
tctgacttct ttccttccgt acgagatctt ctagataccg ccgcagctct gtatcgggat 120
gccttagagt ctcctgagca ttgttcacct caccatactg cactcaggca agcaattctt 180
tgctggggag acttaatgac tctagctacc tgggtgggta ctaatttaga agatccagca 240
tctagggacc tagtagtcag ttatgtcaac actaatgtgg gcctaaagtt cagacaatta 300
ttgtggtttc acatttcttg tctcactttt ggaagagaaa cggttctaga gtatttggtg 360
tcttttggag tgtggattcg cactcctcca gcttatagac caccaaatgc ccctatccta 420
tcaacgcttc cggagactac tgttgttaga cgacgaggca ggtcccctag aagaagaact 480
ccctcgcctc gcagacgaag atctcaatcg ccgcgtcgca gaagatctca atctcgggaa 540
tctcaatgt                                                       549
```

```
<210> 180
<211> 51
<212> DNA
<213> plasmid pKK223
```

```
<400> 180
ttcacacagg aaacagaatt cccggggatc cgtcgacctg cagccaagct t          51
```

```
<210> 181
<211> 38
<212> DNA
<213> plasmid pKK223
```

```
<400> 181
ttcacataag gaggaaaaaa ccatgggatc cgaagctt                        38
```

```
<210> 182
<211> 16
<212> PRT
<213> Hepatitis B virus
```

```
<400> 182
Gly Ile Val Asn Leu Glu Asp Pro Ala Ser Arg Asp Leu Val Val Ser
 1               5                  10                  15
```

```
<210> 183
<211> 17
<212> PRT
<213> Hepatitis B virus

<400> 183
Gly Ile Val Asn Leu Glu Asp Pro Ala Ser Arg Asp Leu Val Val Ser
  1               5                  10                  15

Cys


<210> 184
<211> 4
<212> PRT
<213> Plasmodium falciparum

<400> 184
Asn Ala Asn Pro
  1


<210> 185
<211> 4
<212> PRT
<213> Plasmodium falciparum

<400> 185
Asn Val Asp Pro
  1


<210> 186
<211> 31
<212> DNA
<213> Hepatitis B virus

<400> 186
gcggaattcc atcttccaaa ttaacaccca c                          31
```

(54) Title: MODIFICATION OF HEPATITIS B CORE ANTIGEN

(57) Abstract: A protein is provided comprising hepatitis B core antigen (HBcAg) wherein one or more of the four arginine repeats has been deleted, said protein comprising the C-terminal cysteine of HBcAg. The deleted region may be replaced by an epitope from a protein other than HBcAg, in which case the HBcAg acts as a carrier to present the epitope to the immune system. The chimeric protein is useful in prophylactic and therapeutic vaccination of a host, for example against hepatitis B virus.

MODIFICATION OF HEPATITIS B CORE ANTIGEN

The invention relates to modified forms of the core antigen of hepatitis B virus (HBV) and
to prophylactic and therapeutic vaccines containing the modified antigen.

5

Background to the invention

HBV remains a major healthcare problem throughout both the developed and developing
world. Infection with the virus can result in an acute or chronic disease which in a

10   proportion of cases may lead to hepatocellular carcinoma and death. The virus is double
shelled, and its DNA is protected inside a protein structure called the core antigen
(HBcAg). The core is surrounded by the envelope protein known as the surface or S
antigen (HBsAg).

15   HBcAg is an unusual antigen which can be used as a delivery vehicle for specific peptides
to the immune system. The antigen has been used to present T-helper, B and cytotoxic
lymphocyte (CTL) epitopes from a variety of viral and bacterial pathogens, including
epitopes from the surface antigen of HBV, envelope proteins from hepatitis A and antigens
from hepatitis C virus. For a review see Ulrich et al (1998) Advances in Virus Research

20   50 141-182.

HBcAg is an excellent vehicle for the presentation of epitopes due to the molecular
structure of the protein, which self-assembles into particles. Each particle is generated
from either 180 or 240 copies of a monomeric polypeptide. The polypeptide has 183 or

25   185 amino acids (aa) depending on the subtype of HBV. The monomer, on reaching an
appropriate concentration inside the host cell, forms a particle of approximately 27 nm in
diameter. Structural studies have shown that amino acids within the region from residues
68 to 90 form a spiked structure on the surface of the particle which is known as the e1
loop. Two monomers joined by disulphide bonds link to form a dimer spike, the most

30   exposed amino acid being at position 80 (at the centre of the e1 loop).

MODIFICATION OF HEPATITIS B CORE ANTIGEN

The invention relates to modified forms of the core antigen of hepatitis B virus (HBV) and to prophylactic and therapeutic vaccines containing the modified antigen.

5

Background to the invention

HBV remains a major healthcare problem throughout both the developed and developing world. Infection with the virus can result in an acute or chronic disease which in a
10 proportion of cases may lead to hepatocellular carcinoma and death. The virus is double shelled, and its DNA is protected inside a protein structure called the core antigen (HBcAg). The core is surrounded by the envelope protein known as the surface or S antigen (HBsAg).

15 HBcAg is an unusual antigen which can be used as a delivery vehicle for specific peptides to the immune system. The antigen has been used to present T-helper, B and cytotoxic lymphocyte (CTL) epitopes from a variety of viral and bacterial pathogens, including epitopes from the surface antigen of HBV, envelope proteins from hepatitis A and antigens from hepatitis C virus. For a review see Ulrich et al (1998) Advances in Virus Research
20 50 141-182.

HBcAg is an excellent vehicle for the presentation of epitopes due to the molecular structure of the protein, which self-assembles into particles. Each particle is generated from either 180 or 240 copies of a monomeric polypeptide. The polypeptide has 183 or
25 185 amino acids (aa) depending on the subtype of HBV. The monomer, on reaching an appropriate concentration inside the host cell, forms a particle of approximately 27 nm in diameter. Structural studies have shown that amino acids within the region from residues 68 to 90 form a spiked structure on the surface of the particle which is known as the e1 loop. Two monomers joined by disulphide bonds link to form a dimer spike, the most
30 exposed amino acid being at position 80 (at the centre of the e1 loop).

EP-A-421635 (The Wellcome Foundation Limited) describes modification of the HBV
core gene to allow insertion of foreign epitopes into the e1 loop without altering the
potential of the protein to from particles. Insertion at this site allows maximum exposure
of the inserted epitope on the tip of each spike created by dimers of the protein. As there
5   are approximately 180 (or 240) copies of each monomer per particle, each particle is able
to present 180 (or 240) copies of the epitope of interest.

Thus, HBcAg can be used to generate hybrid particles to be used as prophylactic and
therapeutic vaccines against infectious diseases. However, initial work has identified a
10 high nucleic acid impurity profile due to the inherent nature of the core protein to bind
nucleic acid. The binding of nucleic acid is known to be associated with four arginine
repeats found at the C-terminus of the protein. Removal of these repeats using genetic
tools has been shown to be feasible and results in the production of particles which do not
encapsidate nucleic acid. However, removal of this region appears to reduce the inherent
15 stability of the particle structure.

Summary of the invention

In order to maintain particle stability, whilst overcoming the problem of nucleic acid
20 impurity, the inventors have devised an alternative and novel strategy. The strategy
involves generating a clone in which one or more of the arginine repeats of HBcAg is
removed but in which the C-terminal cysteine is retained. The removal of the arginine
repeats reduces binding of nucleic acid, whilst retention of the C-terminal cysteine allows
the formation of a disulphide bond which in the native structure is important for the
25 formation of a stable particle. The deleted repeat(s) may be replaced with sequences
encoding T-helper, B or CTL epitopes from bacterial or viral pathogens, parasites,
allergens or cancer associated antigens. This is made possible by insertion of a suitable
cloning site in place of the deleted region.

30 Thus, the invention provides a protein comprising HBcAg wherein one or more of the four
arginine repeats is absent and a C-terminal cysteine residue is present. An epitope from a

protein other than HBcAg may be present in place of the absent arginine repeat(s). The protein may be incorporated into a pharmaceutical composition for prophylactic or therapeutic vaccination, for example against HBV.

5   The protein of the invention may comprise a second epitope from a protein other than HBcAg, and the second epitope may be in the e1 loop of HBcAg. By placing a T-helper epitope in the C-terminus and a B-cell epitope in the e1 loop, it is possible to enhance the response to the B-cell epitope through intrastructural T-cell help. In addition, the strategy can be used to double the number of a particular epitope on each particle, by cloning the
10 same sequence into both the e1 loop and the C-terminal region.

Brief description of the drawings

Figure 1: Amino acid sequence of hepatitis B core using the single letter code. The C-
15 terminal sequence (aa135-185) is highlighted to detail the deletion strategy. The 4 arginine (R) repeats are emboldened and underlined for emphasis. Three or four arginine repeat regions are underlined from aa154-178 or aa146-178 respectively. Deletion of the underlined regions with insertion of the *SpeI* restriction site generates constructs encoded by plasmids pTCR$_{154}$ and pTCR$_{146}$ respectively. pTCR$_{154}$ retains the N-terminal arginine
20 repeat, and pTCR$_{146}$ has all 4 arginine repeats deleted.

Figure 2: DNA sequence coding for HBcAg and location and orientation of oligonucleotide primers used for PCR. The position of the *SpeI* restriction site is given for oligos MGR371, MGR369 and MGR370 (see Table 1).
25
Figure 3: DNA and amino acid sequences of pre-S2 and S epitopes inserted into core. Figure 3A shows the sequence of aa20-55 of the pre-S2 region of the HBV ayw subtype. Figure 3B shows the sequence of aa110-147 of the S antigen of the adw subtype. Figure 3C shows the sequence of aa110-157 of the S antigen of the adw subtype.
30
Figure 4: Agarose gel electroporesis of inverse PCR fragments. Lanes 1, 2, 3 and 4 =

-3-

fragments for pTCR$_{146}$, pTCR$_{154}$, pTCSR$_{146}$ and pTCSR$_{154}$ respectively. Lane 5 = size markers. All fragments are of about 5kb as expected.

Figure 5: Immunoblot analysis of expression of core protein in lysates of *E.coli* bacteria
5 transformed with 3' replacement plasmid constructs. All samples express an anti-core antibody reactive protein of various relative molecular weights depending on presence or absence of replacement sequences and size of replacement. Sample order:

Lane 1 = pTCR$_{146}$ *E.coli* HB101

Lane 2 = pTCR$_{146}$/S110-157 *E.coli* HB101

10 Lane 3 = pTCR$_{146}$/S2-2 *E.coli* HB101

Lane 4 = pTCR$_{154}$ *E.coli* HB101

Lane 5 = pTCR$_{154}$/S110-147 *E.coli* HB101

Lane 6 = pTCR$_{154}$/S110-157 *E.coli* HB101

Lane 7 = pTCR$_{154}$/S2-2 *E.coli* HB101

15 Lane 8 = pTCSR$_{146}$ *E.coli* HB101

Lane 9 = pTCSR$_{146}$/S110-157 *E.coli* HB101

Lane 10 = pTCSR$_{146}$/S2-2 *E.coli* HB101.

Figure 6: Immunoblot analysis of expression of S sequence in lysates of bacteria
20 transformed with 3' replacement plasmid constructs. Constructs incorporating the S sequences (lanes 2, 4, 5 and 7) are anti-S antibody reactive. Sample order:

Lane 1 = pTCR$_{146}$ *E.coli* HB101

Lane 2 = pTCR$_{146}$/S110-157 *E.coli* HB101

Lane 3 = pTCR$_{154}$ *E.coli* HB101

25 Lane 4 = pTCR$_{154}$/S110-147 *E.coli* HB101

Lane 5 = pTCR$_{154}$/S110-157 *E.coli* HB101

Lane 6 = pTCSR$_{146}$ *E.coli* HB101

Lane 7 = pTCSR$_{146}$/S110-157 *E.coli* HB101

Lane 8 = Pre-stain marker (Novex).

30

Figure 7: Immunoblot analysis of expression of pre-S2 sequence in lysates of bacteria

transformed with 3' replacement plasmid constructs. Constructs incorporating the pre-S2 sequences (lanes 2, 4 and 6) are pre-S2 antibody reactive. Sample order:

Lane 1 = pTCR$_{146}$ *E.coli* HB101

Lane 2 = pTCR$_{146}$/S2-2 *E.coli* HB101

5   Lane 3 = pTCR$_{154}$ *E.coli* HB101

Lane 4 = pTCR$_{154}$/S2-2 *E.coli* HB101

Lane 5 = pTCSR$_{146}$ *E.coli* HB101

Lane 6 = pTCSR$_{146}$/S2-2 *E.coli* HB101

Lane 7 = Pre-stain marker (Novex).

10

Figure 8: shows averaged anti-HBc responses in mice immunised with various constructs described in the Examples. The titers were calculated as the negative logarithms of the EC50 (effective concentration, 50%) serum dilution on the basis of sigmoidal dose-response curves.

15

Detailed description of the invention

The modifications to the HBcAg sequence

20   As mentioned above, HBcAg is a protein of 183 or 185 amino acids depending on the sub-type of HBV. The extra two amino acids in the 185 form of the protein are located between the first and the second arginine repeats. The sequence of a 185 amino acid form of the protein with a pre-sequence is shown in Figure 1. In Figure 1, the mature HBcAg sequence runs from the Met residue at position 25 to the Cys residue at the extreme C-

25   terminus, with the sequence from residues 1 to 24 being the pre-sequence. The four arginine repeats are located at the following positions:

	Position in mature 183 aa sequence	Position in mature 185 aa sequence (see Figure 1)
first repeat	150-152	150-152
second repeat	157-159	159-161
third repeat	164-167	166-169
5 fourth repeat	172-175	174-177

One or more of the arginine repeats is deleted in the protein of the invention. Thus, it is possible to delete one, two, three or all four of the repeats and to delete the first repeat, the second repeat, the third repeat and/or the fourth repeat. Any combination of the four

10 repeats may be deleted. The first repeat is primarily responsible for RNA binding and the second, third and fourth repeats are primarily responsible for DNA binding, and in a preferred embodiment the first repeat is retained and the second to fourth repeats are deleted in order to specifically reduce DNA binding.

15 A sequence lying between residues 145 and 182 of HBcAg is generally absent in the proteins of the invention, and preferably a sequence lying between residues 150 and 177 is absent. The deleted sequence may comprise the whole of the sequence from residue 145 to residue 182 (or from residue 150 to residue 177) or may comprise only a part of the sequence between those residues. Equally, the deleted sequence may extend on either side

20 of those residues. As used herein, expressions such as "a sequence lying between residues x and y is absent" mean that the sequence which is absent may include residues x and y. Removal of sequence upstream of residue 145 may interfere with the particle-forming ability of the protein and is therefore generally not recommended. In 185 aa forms of HBcAg the deleted sequence may end at residue 184, and in 183 aa forms it may end at

25 residue 182.

The C-terminal cysteine residue in the protein of the invention is typically the natural residue from the C-terminus of HBcAg and is typically preceded by the sequence immediately upstream of the residue in HBcAg. The preceding HBcAg sequence may

comprise from 1 to 7 residues, i.e. 1, 2, 3, 4, 5, 6 or 7 residues. Thus, the C-terminus of
the protein of the invention may have the sequence Gln Cys, Ser Gln Cys, Glu Ser Gln
Cys, Arg Glu Ser Gln Cys, Ser Arg Glu Ser Gln Cys, Gln Ser Arg Glu Ser Gln Cys or Se.
Gln Ser Arg Glu Ser Gln Cys. However, the Cys residue may not be the one from HBcAg;

5 in this case, a protein according to the invention may be constructed by truncating the
HBcAg sequence and replacing the truncated sequence with another sequence including a
Cys residue and optionally an epitope from a protein other than HBcAg. The Cys residue
is typically located at the extreme C-terminal end of the protein of the invention but it may
be a number of amino acid residues from the extreme C-terminal end. For example, it may

10 be from 1 to 20, from 1 to 10 or from 1 to 5 residues from the C-terminus. In any event,
the Cys residue must be able to form a disulphide bond.

The protein of the invention typically comprises the following elements linked in an N-
terminal to C-terminal direction:

15 (i) an N-terminal part of HBcAg which mediates the formation of particles, for example
residues 1 to 144 (or 1 to 146 or 1 to 154), and
(ii) a C-terminal part of HBcAg comprising the C-terminal cysteine;
wherein at least a part of the sequence of HBcAg from between said N-terminal part and
said C-terminal part comprising one or more of the arginine repeats is absent.

20

Where the protein also comprises an epitope from a protein other than HBcAg in place of
the absent arginine repeat(s), the protein typically comprises the following elements linked
in an N- to C-terminal direction:
(i) an N-terminal part of HBcAg which mediates the formation of particles, for example

25 residues 1 to 144 (or 1 to 146 or 1 to 154),
(ii) an epitope from a protein other than HBcAg, and
(iii) a C-terminal part of HBcAg comprising the C-terminal cysteine;
wherein at least part of the sequence of HBcAg between said N-terminal part and said C-
terminal part comprising one or more of the arginine repeats is absent and is replaced by

30 said epitope.

Where the protein comprises an epitope from a protein other than HBcAg in the e1 loop,
the protein typically comprises the following elements linked in an N- to C-terminal
direction:

(i) an N-terminal part of the HBcAg sequence comprising e.g. residues 1 to 67 (or 1 to 74
5 or 1 to 79),

(ii) an epitope from a protein other than HBcAg,

(iii) a second part of the HBcAg sequence comprising e.g. residues 91 to 144 (or 91 to 146,
91 to 154, 86 to 144, 86 to 146, 86 to 154, 80 to 144, 80 to 146 or 80 to 154); and

(iv) a third part of the HBcAg sequence comprising the C-terminal cysteine;

10 wherein at least a part of the sequence of HBcAg from between residue 145 (or 147 or
155) and the C-terminal cysteine comprising one or more of the arginine repeats is absent.

Where the protein of the invention comprises both a first epitope from a protein other than
HBcAg in place of the absent arginine repeat(s) and a second epitope from a protein other

15 than HBcAg in the e1 loop, the protein typically comprises the following elements linked
in an N- to C-terminal direction:.

(i) an N-terminal part of the HBcAg sequence comprising e.g. residues 1 to 67 (or 1 to 74
or 1 to 78);

(ii) an epitope from a protein other than HBcAg,

20 (iii) a second part of the HBcAg sequence comprising e.g. residues 91 to 144 (or 91 to 146,
91 to 154, 86 to 144, 86 to 146, 86 to 154, 80 to 144, 80 to 146 or 80 to 154);

(iv) a further epitope from a protein other than HBcAg, and

(v) a third part of the HBcAg sequence comprising the C-terminal cysteine;

wherein at least a part of the sequence of HBcAg from between residue 145 (or 147 or

25 155) and the C-terminal cysteine comprising one or more of the arginine repeats is absent.

As will be apparent from the above, the inventors specifically contemplate modifying the
HBcAg sequence in a number of ways, including deletion of one or more of the arginine
repeats, insertion of a heterologous epitope in place of the deleted repeat(s) and insertion

30 of a second heterologous in the e1 loop. However, further modification of the HBcAg
sequence is possible. Such further modification may be by way of substitution, insertion,

deletion or extension. The size of an insertion, deletion or extension may, for example, be
from 1 to 200 aa, from 1 to 100 aa or from 1 to 50 aa, from 1 to 20 aa or from 1 to 6 aa in
the sequence of HBcAg. Substitutions may involve a number of amino acids up to, for
example, 1, 2, 5, 10, 20 or 50 amino acids over the length of the HBcAg sequence. The

5 modified protein generally retains the ability to form particles. Substitutions will generally
be conservative and may be made, for example, according to the following Table, in which
amino acids in the same block in the second column and preferably in the same line in the
third column may be substituted for each other.

10 ALIPHATIC	Non-polar	G A P
		I L V
	Polar-uncharged	C S T M
		N Q
	Polar-charged	D E
		K R
AROMATIC		H F W Y

Each part of the HBcAg sequence in the protein of the invention preferably has at least
70% sequence identity to the corresponding sequence of a natural HBcAg protein, such as

15 the protein having the sequence shown in SEQ ID NO: 2. More preferably, the identity is
at least 80%, at least 90%, at least 98%, at least 97% or at least 99%. Methods of
measuring protein sequence (and nucleic acid sequence) identity are well known in the art.
For example, the UWGCG Package provides the BESTFIT programme (Devereux *et al*
(1984) *Nucleic Acids Research* 12, p.387-395). Similarly, the PILEUP and BLAST

20 algorithms can be used to line up sequences (for example as described in Altschul S. F.
(1993) *J. Mol. Evol.* 36:290-300 and Altschul, S. F. *et al* (1990) *J. Mol. Biol.* 215:403-10).

The protein of the invention may self-assemble into particles which may closely resemble
the particles formed by native HBcAg. The particles may be from 20 to 40 nm in

25 diameter, but are preferably about 27 nm in diameter (which is the size of native HBcAg

particles). They contain no detectable or reduced amounts of nucleic acid (DNA and RNA) compared to particles of native HBcAg. They may contain from 160 to 260 monomers of the protein of the invention, but preferably they contain approximately 180 or approximately 240 monomers (which are the numbers of monomers in native HBcAg

5 particles).

Determination of the particulate nature of a protein according to the invention may be carried out by size exclusion chromatography and/or electron microscopy. Determination of the DNA content of the particles may be carried out by agarose gel electrophoresis or

10 spectrophotometry. A method adapted from Birnbaum and Nasal (1990, J. Virology 64 3319-3330) may be used. The protein may be digested with Proteinase K and the nucleic acid extracted using a commercial DNA recovery kit (e.g. Qiagen, QIAquick™ PCR Purification Kit). Purified DNA may be visualised using a high sensitivity DNA stain (e.g. Novex, SYBER Green I™) in a 1.5% agarose gel, following electrophoresis. The DNA

15 product obtained following extraction may be quantified using the optical density (OD) 260nm:280nm ratio according to Sambrook et al. (1989, Molecular cloning - A laboratory manual, second edition, published by Cold Spring Harbor Laboratory Press), for example using a Pharmacia Biotech Ultraspec 2000™.

20 **The epitopes**

As a general rule, epitopes inserted into the protein of the invention should not prevent the folding of HBcAg or its self-assembly into particles. In addition, for improved immunogenicity, B-cell epitopes should be displayed on the surface of the particle. T-cell

25 epitopes do not need to be displayed on the surface of the particle for optimal presentation.

There are three preferred regions for insertion of the epitopes, namely the C-terminus in place of deleted arginine repeat(s), the e1 loop and the N-terminus. These three regions all tolerate well insertion of foreign sequences. When an epitope is placed in the e1 loop of

30 HBcAg, it may be inserted in the sequence of amino acid residues 68 to 90, 69 to 90, 71 to 90, 75 to 85 or 78 to 83. Most preferred is to insert the epitope between residues 79 and 80

or 80 and 81. HBcAg residues from the e1 loop may be deleted in proteins of the invention, so that the inserted epitope may replace all or part of the sequence of the loop.

A heterologous epitope present in a protein of the invention may be a B-cell epitope or a

5   T-cell epitope. In the case that an epitope is a T-cell epitope it may be a T-helper (Th) cell epitope (either a Th1 or Th2 epitope) or a cytotoxic lymphocyte (CTL) epitope.

The protein of the invention may contain more than one heterologous epitope, for example up to 2, 3, 5 or 8 heterologous epitopes, and in this case each epitope may be present in the

10 same site or at different sites in HBcAg. In a preferred embodiment of the invention, one of the epitopes is a T-helper cell epitope and another is a B-cell or a CTL epitope. The presence of the T-helper cell epitope enhances the immune response against the B-cell or CTL epitope. Where there are two or more heterologous epitopes in the protein of the invention, they may be from the same organism or the same protein. Indeed, the epitopes

15 may be the same; this allows a doubling or further multiplication of the number of the epitope presented on the particles.

The size of the sequence comprising an epitope inserted in the protein of the invention can vary between broad limits, but will generally be from 6 to 120 aa, for example from 6 to

20 80 aa or 6 to 40 aa. The epitope may be conformational or linear.

The choice of epitope depends on the disease that it is wished to vaccinate against. Typically, the epitope is from a pathogen, such as a virus, a bacterium or a protozoan, but it may also be from a cancer associated antigen or an allergen. Examples of pathogens

25 whose epitopes may be inserted include hepatitis A virus (HAV), HBV, hepatitis C virus (HCV), influenza virus, foot-and-mouth disease virus, poliovirus, herpes simplex virus, rabies virus, feline leukemia virus, human immunodeficiency virus type 1 (HIV1), human immunodeficiency virus type 2 (HIV2), simian immunodeficiency virus (SIV), human rhinovirus, dengue virus, yellow fever virus, human papilloma virus, *Plasmodium*

30 *falciparum* (a cause of malaria) and bacteria such as *Mycobacteria, Bordetella, Salmonella, Escherichia, Vibrio, Haemophilus, Neisseria, Yersinia* and *Brucella.*

-11-

Specifically, the bacterium may be *Mycobacterium tuberculosis* - the cause of
tuberculosis; *Bordetella pertussis* or *Bordetella parapertussis* - causes of whooping cougj
Salmonella typhimurium - the cause of salmonellosis in several animal species; *Salmonel..*
typhi - the cause of human typhoid; *Salmonella enteritidis* - a cause of food poisoning in

5   humans; *Salmonella choleraesuis* - a cause of salmonellosis in pigs; *Salmonella dublin* - a
cause of both a systemic and diarrhoeal disease in cattle, especially of new-born calves;
Escherichia coli - a cause of food poisoning in humans; *Haemophilus influenzae* - a cause
of meningitis; *Neisseria gonorrhoeae* - a cause of gonnorrhoeae; *Yersinia enterocolitica* -
the cause of a spectrum of diseases in humans ranging from gastroenteritis to fatal

10 septicemic disease; *Brucella abortus* - a cause of abortion and infertility in cattle and a
condition known as undulant fever in humans; or *Clostridium difficile* - a cause of
pseudomembranous colitis.

Examples of antigens whose epitopes may be inserted are the pre-S1, pre-S2 and S

15 antigens of HBV; the HAV surface antigens; the HCV surface antigens, core protein and
NS3 protein; the HIV antigens gp120, gp160, gag, pol, Nef, Tat and Ref; the malaria
antigens such as the circumpsorozoite proteins; the influenza antigens HA, NP and NA;
the herpes virus antigens EBV gp340, EBV gp85, HSV gB, HSV gD, HSV gH and HSV
early protein; the human papilloma virus antigens E4, E6 and E7; the cancer antigens

20 carcinoembryonic antigen (CEA), P53, ras and myc; the pertactin antigen from *Bordetella*
pertussis; and house dust mite allergen.

The invention is particularly suited to prophylactic or therapeutic vaccination against HBV
since the carrier protein HBcAg is from HBV, and epitopes from the pre-S1, pre-S2 and S

25 regions of HBV are particularly preferred. A pre-S1, pre-S2 or S insert is typically at least
6 amino acids in length, for example from 6 to 120 aa, 8 to 80 aa or 10 to 40 aa. The insert
may include, for example, the residues at pre-S1 positions 1-9, 10-19, 20-29, 30-39, 40-49,
50-59, 60-69, 70-79, 80-89, 90-99, 100-109 or 110-119 or the residues at pre-S2 positions
120-129, 130-139, 140-149, 150-159, 160-169 or 170-174. Particularly preferred

30 fragments are those corresponding to pre-S1 residues 20-47 and pre-S2 residues 139-174.
Pre-S1 residues 21-28 correspond to a human T-cell epitope. Also preferred are fragments

-12-

corresponding to S residues 110-147 and 110-157 (counting the first residue of the S sequence as residue 1).

Making the proteins of the invention

5

The proteins of the invention are generally made by recombinant DNA technology. The invention includes a nucleic acid molecule (e.g. DNA or RNA) encoding a protein of the invention, such as an expression vector.

10 The nucleic acid molecule may encode a protein in which one or more of the arginine repeats has been deleted and replaced with a restriction enzyme site unique to the nucleic acid molecule, such as an XbaI site. The nucleic acid molecule may also contain a unique restriction enzyme site in the sequence encoding the e1 loop and/or in the N-terminus. The unique restriction enzyme sites allow sequences encoding epitopes to be inserted into the
15 nucleic acid molecule, for example in place of the deleted arginine repeat(s) or in the e1 loop.

A protein of the invention may be produced by culturing a host cell containing a nucleic molecule encoding the protein under conditions in which the protein is expressed, and
20 recovering the protein. Suitable host cells include bacteria such as *E. coli*, yeast, mammalian cell lines and other eukaryotic cell lines, for example insect Sf9 cells.

The vectors constituting nucleic acid molecules according to the invention may be, for example, plasmid or virus vectors. They may contain an origin of replication, a promoter
25 for the expression of the sequence encoding the protein, a regulator of the promoter such as an enhancer, a transcription stop signal, a translation start signal and/or a translation stop signal. The vectors may also contain one or more selectable marker genes, for example an ampicillin resistance gene in the case of a bacterial plasmid or a neomycin resistance gene for a mammalian vector. Vectors may be used *in vitro*, for example for the production of
30 RNA or used to transform or transfect a host cell. The vector may also be adapted to be used *in vivo*, for example in a method of gene therapy or DNA vaccination.

Promoters, enhancers and other expression regulation signals may be selected to be compatible with the host cell for which the expression vector is designed. For example, prokaryotic promoters may be used, in particular those such as the *trc* promoter suitable for use in *E. coli* strains (such as *E. coli* HB101). A promoter whose activity is induced in

5   response to a change in the surrounding environment, such as anaerobic conditions, may be used. Preferably an *htrA* or *nirB* promoter may be used. These promoters may be used in particular to express the protein in an attenuated bacterium, for example for use as a vaccine. When expression of the protein of the invention is carried out in mammalian cells, either *in vitro* or *in vivo*, mammalian promoters may be used. Tissue-specific

10   promoters, for example hepatocyte cell-specific promoters, may also be used. Viral promoters may also be used, for example the Moloney murine leukaemia virus long terminal repeat (MMLV LTR), the rous sarcoma virus (RSV) LTR promoter, the SV40 promoter, the human cytomegalovirus (CMV) IE promoter, herpes simplex virus promoters and adenovirus promoters. All these promoters are readily available in the art.

15

A protein according to the invention may be purified using conventional techniques for purifying proteins. The protein may, for example, be provided in purified, pure or isolated form. For use in a vaccine, the protein must generally be provided at a high level of purity, for example at a level at which it constitutes more than 80%, more than 90%, more than

20   95% or more than 98% of the protein in the preparation. However, it may be desirable to mix the protein with other proteins in the final vaccine formulation, for example other proteins comprising pre-S1, pre-S2 or S sequence of HBV. The protein is preferably substantially free from nucleic acid (DNA and RNA).

25 **Vaccines**

The primary use of the proteins of the invention is as therapeutic or prophylactic vaccines. The invention includes a pharmaceutical composition (e.g. a vaccine composition) comprising a protein of the invention, a particle of the invention or a nucleic acid molecule

30   of the invention and a pharmaceutically acceptable carrier or diluent.

The principle behind prophylactic vaccination is to induce an immune response in a host so as to generate an immunological memory in the host. This means that, when the host is exposed to the virulent pathogen, it mounts an effective (protective) immune response, i.e an immune response which inactivates and/or kills the pathogen. The invention could form the basis of a prophylactic vaccine against a range of diseases, such as HBV, HAV, HCV, influenza, foot-and-mouth disease, polio, herpes, rabies, AIDS, dengue fever, yellow fever, malaria, tuberculosis, whooping cough, salmonellosis, typhoid, food poisoning, diarrhoea, meningitis and gonnorrhoeae. The epitopes in the protein of the invention are chosen so as to be appropriate for the disease against which the vaccine is intended to provide protection.

The principle behind therapeutic vaccination is to stimulate the immune system of the host to alleviate or eradicate a disease or condition. There are a number of diseases and conditions which may be susceptible to therapeutic vaccination, such as chronic viral diseases including chronic HBV and chronic HCV, cancer, and allergies such as asthma, atopy, eczema, rhinitis and food allergies.

Chronic viral diseases arise when the immune system of an infected host fails to eliminate the virus, allowing the virus to persist in the host for a long period of time. The invention may be used to induce the immune system of the chronically infected individual so as to eliminate the virus. For example, it is believed that patients with chronic hepatitis have an inadequate T-cell response, and that stimulation of an appropriate T-cell response can eliminate the virus. Thus, in order to treat viral hepatitis using the invention, T-cell epitopes may be inserted into the protein of the invention, such as T-cell epitopes from the pre-S1 and pre-S2 regions of HBV.

Similarly, in the case of cancer, it is believed that enhancement of the T-cell response to tumour antigens may help the immune system to destroy the tumour. It is believed that allergic diseases are caused at least in part by an unbalanced T-cell response in which an inflammatory Th2 responses dominates over an antagonistic Th1 response, and that

allergies may therefore be treated by enhancing the Th1 response. This can be achieved according to the invention by using a protein which stimulates a Th1 response.

More than one protein according to the invention may be administered to a patient.

5 Furthermore, a protein according to the invention may be used in combination with one or more other compositions. For example, in the treatment of chronic HBV a protein according to the invention may be used in combination with interferon gamma, Lamivudine™, or another immunotherapeutic agent such as Hepacare™ (formerly known as Hepagene™). The protein according to the invention and the other composition may be

10 administered simultaneously or sequentially.

Suitable carriers and diluents for inclusion in pharmaceutical compositions of the invention are isotonic saline solutions, for example phosphate-buffered saline. The composition will normally include an adjuvant, such as aluminium hydroxide. The

15 composition may be formulated for parenteral, intramuscular, intravenous, intranasal, subcutaneous or transdermal administration. The composition comprises the protein, particles or nucleic acid in a prophylactically or therapeutically effective amount. Typically, the protein or particles are administered at a dose of from 0.01 to 30 μg/kg body weight, preferably from 0.1 to 10 μg/kg, more preferably from 0.1 to 1 μg/kg body weight.

20 The nucleic acid of the invention may be administered directly as a naked nucleic acid construct using techniques known in the art or using vectors known in the art. The amount of nucleic acid administered is typically in the range of from 1 μg to 10 mg, preferably from 100 μg to 1 mg. The vaccine may be given in a single dose schedule or a multiple dose schedule. The routes of administration and doses given above are intended only as a

25 guide, and the route and dose may ultimately be at the discretion of the physician.

30

Experimental Section

Experiment 1

5 1.   **Materials and Methods**

New plasmid constructs were generated by inverse PCR so that three or four C-terminal
arginine repeat regions were deleted and a *Spe*I restriction site was introduced to allow
insertion of replacement sequences coding for B and T cell epitopes (Fig. 1).

10

The plasmid templates for the inverse PCR were p*trc*/core and p*trc*/core-S1 which encode
respectively for non-hybrid hepatitis B core and hybrid hepatitis B core containing amino
acids 20-47 of the pre-S1 sequence of hepatitis B surface protein inserted between amino
acids 79 and 80 of the immunodominant e1 loop. Three oligonucleotide primers (Table 1

15   and Fig. 2) were used for the PCR reaction. These primers introduce a unique *Spe*I
restriction site in the PCR fragments. The primers were also designed to generate new
fragments that were truncated at residues 146 or 154 but maintained 7 residues of the C-
terminus including the terminal cysteine at position 185 which is thought to be important
for maintaining particle stability by formation of disulphide bonds (Fig. 1).

20

        1.1   Construction of parental truncated plasmids

Using primers MGR371/370 or MGR369/370 (Table 1 and Figure 2), inverse PCR
fragments are generated from plasmid templates of p*trc*/core or p*trc*/core-S1. This

25   procedure removes 69 nucleotides (encoding for 23 amino acids (aa155-177)) and 93
nucleotides (encoding for 31 amino acids (146-177)) respectively. The PCR fragments
sizes were confirmed by analysis on agarose gels and then digested with *Spe*I restriction
endonuclease followed by purification on agarose gels and self-ligation to generate
plasmids pTCR$_{146}$, pTCR$_{154}$ and pTCSR$_{146}$ and pTCSR$_{154}$. pTCR plasmids are derived

30   from the p*trc*/core template and pTCSR plasmids are derived from the p*trc*/core-S1
templates. The 146 and 154 numbering denotes the amino acid number at the truncation

point. The four parental truncated plasmids were used to transform *E.coli* HB101 cells and positive colonies were tested by diagnostic PCR using oligonucleotide primers MGR61/MGR168. Core protein expression was confirmed by immunoblotting of bacterial cell lysates using a mouse anti-core antibody.

5

1.2   Subcloning of replacement sequences into truncated parental plasmids

Three sequences have been subcloned into the 3' end of the truncated parental plasmids described in section 1.1. These include sequences encoding for amino acids 110-147 and

10   110-157 of the small hepatitis B surface protein, and aa20-55 of the S2 region of the middle hepatitis B surface protein (Figure 3).

For insertion of the 110-157 sequence (plus 2 amino acids resulting from the *Nhe*I restriction site) oligonucleotide primers MR245-247 (Table 1B) were used to generate a

15   PCR fragment of 147 nucleotides using pMBdSRE/17 as template (Figure 3). This plasmid encodes for the small hepatitis B surface protein (*adw* subtype) for expression in mammalian cells using the mouse metallothionine promoter.

For insertion of the 110-147 sequence (plus 2 amino acids from the *Nhe*I site)

20   oligonucleotide primers MGR247/264 (Table 1B) were used to generate a PCR fragment of 120 nucleotides using pMBdSRE/17 as template (Figure 3).

For insertion of the 20-55 sequence (plus 2 residues from the *Nhe*I site) of pre-S2, oligonucleotide primers MGR243/249 (Table 1B) were used to generate a PCR fragment

25   of 114 nucleotides using pMByS2R/8 as template (Figure 3). This plasmid encodes for the middle hepatitis B surface protein (*ayw* subtype) under control of the metallothionine promoter for mammalian cell expression.

The PCR fragments were digested with *Nhe*I restriction endonuclease and purified on

30   agarose gels. The purified fragments were then ligated with *Spe*I digested, phosphatase treated parental plasmids (section 1.1). *E.coli* HB101 cells were then transformed with the

resulting plasmids and positive colonies tested by diagnostic PCR using oligonucleotide primers MGR61/168, immunoblotting with antibodies specific for the insert and partial DNA sequencing of the inserts.

5   2.   Results

2.1   Confirmation of inverse PCR fragment generation

Inverse PCR fragments for $pTCR_{146}$, $pTCR_{154}$, $pTCSR_{146}$ and $pTCSR_{154}$ were analysed by
10   separation on 1% agarose gels (Figure 4). The PCR fragments were found to be of the appropriate size (approx. 5.2kb) and were confirmed to be correct by diagnostic PCR (not shown). Immunoblot analysis showed that the parental constructs and those containing the inserted sequences expressed the core protein that was reactive to an anti-core antibody (Figure 5). Further, confirmation of protein expression of the inserted sequences was
15   shown by immunoblotting using anti-S (Figure 6) and anti-pre-S2 antibodies (Figure 7).

Table 1. Oligonucleotide primers used for inverse and diagnostic PCR

Table 1A

Oligos	5'-3' sequence	ptrc/core
MGR61	CTGCACTCAGGCAAGCCATT	230bp-249bp
MGR62	GCCGAGGCAGGTCCCCTAGA	530bp-549bp
MGR168	GAAAATCTTCTCGGATCCGC	from vector (pKK233.2)
MGR282	AGAGATCTCCATGGATTCAG	-10bp-10bp
MGR280	GTGGCTTTGGGGCCATGGACA	60bp-79bp
MGR369	AGGACTAGTGCCTCGGCCCCGTCGTCT	520bp-546bp
MGR370	AGAACTAGTCAATCTAGGGAATCTCAA	598bp-624bp
MGR371	TCTTCTAACACTAGTAGTTTCCGG	502bp-525bp

Bold denotes *SpeI* sites

30

Table 1B

Oligos	5'-3' sequence	gene and loca
MGR245	CAGCTAGCGCAATTTCCATCCGTA	HBsAg 147aa
MGR247	GTTTGTGCTAGCATTCCAGGAACA	HBsAg 110aa
MGR264	CCATAGGTTGCTAGCGAAAGCCCA	HBsAg 157aa
MGR243	TTGCTAGCGTTCAGCGCAGGGTCC	Pre-S2 20aa
MGR249	GTGAGAGCTAGCTATTTCCCTGCT	Pre-S2 55aa

Experiment 2

Summary

Full-length and C-terminally truncated hepatitis B core antigen (HBc) derivatives, which carried long foreign amino acid insertions at position 144, were constructed. HBV preS1, preS2, and HIV-1 Gag fragments of 50-100 amino acids in length were used as such insertions, and the appropriate recombinant genes were expressed in *E.coli* cells. The appropriate chimeric HBc and HBcΔ derivatives were purified and examined antigenically and immunogenically. Subclass analysis of the induced anti-HBc immune response in mice showed that the Ig ratio of IgG1, IgG2a, and IgG2b antibodies was restored from the IgG1>IgG2a≥IgG2b pattern, which is typical for C-terminally truncated HBcΔ derivatives, to IgG2a≥IgG2b≥IgG1, which is typical for full-length HBc derivatives, after immunisation with C-terminally truncated HBcΔ derivatives which carried long C-terminal additions of 50-100 amino acids in length.

Materials and Methods

Bacterial Strains

E.coli strains RR1 (F, *hsd*S20 (r⁻ᵦ, m⁻ᵦ), *recA⁺, ara*-14, *proA2, lacY1, galK2, rpsL20* (Sm'), *xyl*-5, *mtl*-1, *supE44*, λ⁻), and K802 (*hsdR, gal, met, supE, mcrA, mcrB*) were used for selection and expression of chimeric genes, respectively.

Animals

BALB/C (H-2⁴) female mice were used approximately 7-10 weeks old, weight 20
mg. New Zealand white strain female rabbits were used for obtaining polyclonal
antibodies.

5

Construction of HBc Derivatives

Vectors based on plasmids pHBc3 and pHBc16-15. Vector pHBc3 was constructed by putting
the HBc gene under the control of the tandem repeat of *E.coli trp* promoters. Vector pHBc16-
10   15 was constructed by insertion of an oligonucleotide linker carrying *Cla* I/*Eco RV* restriction
sites into position 144 of the HBc gene.

Construction of chimeric HBc derivatives. The structure of the HBc and HBcΔ derivatives is
shown in Table 2. The recombinant genes were constructed by insertion of the appropriate
15   HBV preS1, preS2, and HIV-1 gag fragments into the Cla I site of the pHBc16-15 vector, with
or without in-frame junction to the C-terminal part of the HBc gene.

Purification of Chimeric HBc Derivatives

E.coli cells were grown overnight on a rotary shaker at 37⁰C in 750 ml flasks containing 300
20   ml of M9 minimal medium supplemented with 1% casamino acids (Difco Laboratories,
Sparks, USA) and 0.2% glucose. An optical density OD_{540} of 2-5 was usually reached.
Generally, cells were pelleted and lysed by 30 min incubation on ice in lysis buffer containing
50 mM Tris-HCl (pH 8.0), 5 mM EDTA, 50 µg/ml PMSF, 2 mg/ml lysozyme and then
ultrasonicated 3 times for 15 s at 22 kHz. Lysates were then adjusted to 10 mM $MgCl_2$, and
25   20 µg/ml DNAase. After low speed centrifugation, proteins were precipitated from the
supernatant with ammonium sulfate at 33% saturation for 1-2 h at 4⁰C. Pellets were
resuspended in a standard PBS buffer containing 0.1% Triton X-100™, and 5 ml of the
solutions were loaded on a Sepharose CL4B™ column (2.5 x 85 cm) and eluted with PBS
buffer without Triton X-100. The presence of HBc polypeptides in fractions was tested by
30   PAGE. Positive fractions were pooled and concentrated by ammonium sulfate precipitation
at 33% saturation for 20 h at 4⁰C. Pellets were resuspended in PBS, or in Tris-saline buffer,

10 mM Tris-HCl (pH7.5), 150 mM NaCl, to a final concentration of about 5-20 mg/ml, dialyzed overnight against 2000 volumes of the same buffer and stored at -70°C or at -20°⌐ in 50% glycerol.

5 *Polyacrylamide Gel Electrophoresis and Western Blotting*

For PAGE analysis, bacteria were pelleted, suspended in SDS-gel electrophoresis sample buffer containing 2% SDS and 2% 2-mercaptoethanol and lysed by heating at 100°C for 5 min. The proteins were separated by Laemmli's polyacrylamide gel electrophoresis (PAGE) in a slab gel (150x150x0.75 mm) apparatus with a gradient 12-18% running gel and

10 a 4% stacking gel. Western blotting was performed in general as described by Towbin et al (1979) in Proc. Natl. Acad. Sci. USA 76 4350-4354. Nitrocellulose sheets (0.2 μ, Millipore, Bedford, USA) were incubated with anti-HBc antibodies and anti-preS1 antibody in dilutions of 1:100 to 1:1000 overnight and then with anti-mouse IgG peroxidase conjugate (1:1000) for 1-2 h at room temperature. The reaction was developed with 3,3'-diaminobenzidine. In

15 parallel, gels were silver-stained according to Ohsawa and Ebata (1983) Anal. Biochem. 135 409-415.

Immunisations

Mice (five per group) were immunised at day 0 intraperitoneally with 0.02 mg of chimeric

20 particles in complete Freund's adjuvant (CFA, Difco) followed by two booster immunisations in Freund's incomplete adjuvant (IFA, Difco) given at days 10 (0.01 mg intraperitoneally) and 24 (0.01 mg intraperitoneally and 0.01 mg subcutaneously). Sera obtained on day 32 were analysed by ELISA for reactivity with HBc particles.

25 *ELISA*

For the ELISA, recombinant HBc particles were coated onto 96-well microtiter plates by air-drying in a chemical hood overnight. Wells were blocked with 0.5% BSA in PBS for 1 h, incubated with serial dilutions of the various antibodies for 1 h at 37°C and processed with the appropriate second antibodies conjugated to horse radish peroxidase (Sigma) according to the

30 protocols of the manufacturers. Plates were washed 5 times between incubations with 0.05% Tween-20™ in PBS, and 5 times with distilled water to remove Tween-20. Optical

absorbances were measured at 492 nm in an automatic Immunoscan MS™ reader. The titres were calculated as the negative logarithms of the EC50 (effective concentration, 50%) seru dilution on the basis of sigmoidal dose-response curves. GraphPad Prism® version 3.ᵔᴗ software was used in the mean titre calculations.

5

Results

Immunogenicity of Recombinant Proteins. To measure the immunogenicity of HBc carrier and inserted preS1, preS2, and Gag sequences, individual mice sera were repeatedly tested by

10 direct ELISA using recombinant HBcAg and synthetic preS1, preS2, and HIV-1 p24 peptides on solid support. Immunisation with chimeric particles induced high levels of anti-HBc and relatively low levels of anti-insertion antibodies (not shown).

Induction of Different Immunoglobulin Subclasses by Chimeric HBcΔ-preS1 (20-47) Particles

15 In order to average obtained immunisation data and to make them more informative for comparative subclass analysis of induced immunoglobulins, we calculated mean titres for each group of immunised animals as the negative logarithms of the EC50 (effective concentration, 50%) serum dilution on the basis of sigmoidal dose-response curves (GraphPad Prism® version 3.02). These data on the anti-HBc response of immunised mice, which allow direct

20 comparison of averaged titres, are given in Fig. 8.

The data presented in Fig. 8 show that the wild type HBcAg induces anti-HBc response with the immunoglobulin subclass distribution IgG2a≥IgG2b>IgG1, whilst the immune response to the C-terminally truncated HBcΔ structure T31 presents the IgG1>IgG2b≥IgG2a subclass

25 distribution pattern. The full-length HBc derivative 10-62, which carries a 50 aa long preS1 insertion, shows a subclass distribution analogous to that of the full-length HBc vector. Moreover, replacement of the C-terminus of the HBc molecule by a long foreign insertion (50 amino acids of the preS1 sequence) in the HBc derivative 10-140 makes the subclass distribution of the anti-HBc antibodies rather similar to that induced by the full-length HBc

30 structure (Fig. 8). The HBcΔ derivative 48-2 with a 100 aa long insertion of HIV-1 Gag

-23-

occupies an intermediate position in this sense between wild type HBcAg and C-terminally truncated HBcΔ T31 structures.

Table 2. Structure of HBc derivatives with C-terminal insertions. Amino acids appearing at the HBc and insertion sequence junctions are shown in lowercase.

Full-length HBc derivatives

Constr uct	Inser tion	Sequence							
HBc 10-62	preS 131-80	HBc 144 P krsiskrsis	<u>preS1</u> 31 40 50 60 70 80 DPAFRANTANPDWDFNPNKDTWPDANKVGAGAFGLGFTPPHGGLLGWSPQ s						HBc 145 E..
HBc 9-87	preS 21-54	HBc 144 P krsi	<u>preS2</u> 1. 10 20 30 40 50 QAMQWNSTTFHQTLQDPRVRGLYFPAGGSSSGTVNPVPTTVSPISSIFSRIGDPAL ks						HBc 145 E..

C-terminally truncated HBc derivatives

Constr uct	Inser tion	Sequence							
HBcΔ 10-140	preS 131-79	HBc 144 P krsiskrsis	<u>preS1</u> 31 40 50 60 70 79 DPAFRANTANPDWDFNPNKDTWPDANKVGAGAFGLGFTPPHGGLLGWSP hdigdycc						
HBcΔ 9-142	preS 21-55	HBc 144 P krsi	<u>preS2</u> 1 10 20 30 40 50 55 QAMQWNSTTFHQTLQDPRVRGLYFPAGGSSSGTVNPVPTTVSPISSIFSRIGDPALN gdycc						
HBcΔ 48-2	HIV p55 121-210	144 p17 <u>p24</u> 1 P ns DTGHSSQVSQNYPIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQDLNTM 56 78 LNTVGGHQAAMQMLKETINEEAA agmqasla							55

CLAIMS

1. A protein comprising hepatitis B core antigen (HBcAg) wherein one or more of the four arginine repeats is absent and a C-terminal cysteine residue is present.

5

2. A protein according to claim 1 wherein a first epitope from a protein other than HBcAg is present in place of the absent arginine repeat(s).

3. A protein according to claim 1 or 2 wherein the first arginine repeat is present and

10   the second to fourth arginine repeats are absent.

4. A protein according to any one of the preceding claims wherein a sequence lying between residues 145 and 182 of HBcAg is absent.

15 5. A protein according to any one of the preceding claims wherein a sequence lying between residues 150 and 177 of HBcAg is absent.

6. A protein according to any one of the preceding claims which comprises a second epitope from a protein other than HBcAg, the second epitope being in the e1 loop.

20

7. A protein according to claim 6 wherein the second epitope is a B-cell epitope.

8. A protein according to any one of claims 2 to 7 wherein the first epitope is a T-cell epitope.

25

9. A protein according to claim 8 wherein the first epitope is a T-helper cell epitope and the second epitope is a B-cell epitope.

10. A protein according to claim 6 which comprises said first and second epitopes

30   wherein the epitopes are the same.

-26-

11.   A protein according to any one of claims 2 to 10 wherein the first and/or the second
       epitope is from hepatitis B virus (HBV).

12.   A protein according to claim 11 wherein the first and/or the second epitope is from
5   the pre-S1, pre-S2 or S region of HBV.

13.   A protein according to claim 1 comprising the following elements linked in an N-
       terminal to C-terminal direction:
       (i) an N-terminal part of HBcAg which mediates the formation of particles, and
10   (ii) a C-terminal part of HBcAg comprising the C-terminal cysteine; .
       wherein at least a part of the sequence of HBcAg from between said N-terminal
       part and said C-terminal part comprising one or more of the arginine repeats is
       absent.

15 14.   A protein according to claim 1 comprising the following elements linked in an N-
       to C-terminal direction:
       (i) an N-terminal part of HBcAg which mediates the formation of particles,
       (ii) an epitope from a protein other than HBcAg, and
       (iii) a C-terminal part of HBcAg comprising the C-terminal cysteine;
20   wherein at least a part of the sequence of HBcAg between said N-terminal part and
       said C-terminal part comprising one or more of the arginine repeats is absent and is
       replaced by said epitope.

15.   A protein according to claim 1 comprising the following elements linked in an N-
25   to C-terminal direction:
       (i) an N-terminal part of the HBcAg sequence comprising residues 1 to 67,
       (ii) an epitope from a protein other than HBcAg,
       (iii) a second part of the HBcAg sequence comprising residues 91 to 144, and
       (iv) a third part of the HBcAg sequence comprising the C-terminal cysteine;
30   wherein at least a part of the sequence of HBcAg from between residue 145 and the
       C-terminal cysteine comprising one or more of the arginine repeats is absent.

-27-

16.   A protein according to claim 1 comprising the following elements linked in an N-
to C-terminal direction:

(i) an N-terminal part of the HBcAg sequence comprising residues 1 to 67;

(ii) an epitope from a protein other than HBcAg,

(iii) a second part of the HBcAg sequence comprising residues 91 to 144;

(iv) a further epitope from a protein other than HBcAg;

(v) a third part of the HBcAg sequence comprising the C-terminal cysteine;

wherein at least a part of the sequence of HBcAg from between residue 145 and the
C-terminal cysteine comprising one or more of the arginine repeats is absent.

17.   A particle comprising multiple copies of a protein as claimed in any one of the
preceding claims.

18.   A nucleic acid molecule encoding a protein as claimed in any one of claims 1 to 16.

19.   A nucleic acid molecule according to claim 18 which is an expression vector.

20.   A host cell transformed or transfected with a nucleic acid molecule as claimed in
claim 18 or 19.

21.   A process for producing a protein as claimed in any one of claims 1 to 16, which
process comprises culturing a host cell containing a nucleic acid molecule which
encodes the protein under conditions in which the protein is expressed, and
recovering the protein.

22.   A nucleic acid molecule encoding a protein as claimed in claim 1 wherein the
sequence encoding one or more of the four arginine repeats of HBcAg is deleted
and replaced with a restriction enzyme site unique to the nucleic acid molecule.

23.   A pharmaceutical composition comprising a protein as claimed in any one of
claims 1 to 16, a particle as claimed in claim 17 or a nucleic acid molecule as

claimed in claim 18 or 19 and a pharmaceutically acceptable carrier or diluent.

24.   A protein according to any one of claims 1 to 16, a particle according to claim 17 ..
a nucleic acid molecule according to claim 18 or 19 for use in a method of

5   prophylactic or therapeutic vaccination of the human or animal body.

25.   A protein, particle or nucleic acid molecule according to claim 24 for use in a
method of prophylactic or therapeutic vaccination of the human or animal body
against HBV.

10

26.   Use of a protein according to any one of claims 1 to 16, a particle according to
claim 17 or a nucleic acid molecule according to claim 18 or 19 for the manufacture
of a medicament for prophylactic or therapeutic vaccination of the human or animal
body against HBV.

15

27.   A method of vaccination or therapy of a subject, which method comprises
administering to the subject a protein as claimed in any one of claims 1 to 16, a
particle as claimed in claim 17 or a nucleic acid molecule as claimed in claimed 18
or 19.

20

Fig.1.

MDSDNPASTTNKDKDPRALGWLWGMDIDPYKEFGATVELLSFLPSDFFPSVR
DLLDTASALYREALESPEHCSPHHTALRQAILCWGELMTLATWVGNNLEDPAS
RDLVVNYVNTNMGLKIRQLLWFHISCLTFGRETVLEYLVSFGVWIRTPPAYRP

140⟶ 150⟶ 160⟶ 170⟶ 180⟶ 185⟶

PNAPILSTLPETTVVRRRGRSPRRRTPSPRRRRSQSPRRRRSQSRESQC

PNAPILSTLPETTVVRRRGTSQSRESQC   aa sequences for pTCR₁₅₄
                    *Spe*1

PNAPILSTLPETSQSRESQC   aa sequences for pTCR₁₄₆
             *Spe*1

Fig.2.

AGAGATCTCC

    10   20   30   40   50
ATGGATTCAGATAATCCAGCTAGTACTACCAACAAAGATAAGGATCCGCG

MGR282

    60   70   80   90   100
CGCCCTTGGGTGGCTTTGGGGCATGGACATTGACCCTTATAAAGAATTTG

MGR280

   110   120   130   140   150
GAGCTACTGTGGAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCC

   160   170   180   190   200
GTCAGAGATCTTCTAGACACCGCCTCAGCTCTGTATCGAGAAGCCTTAGA

   210   220   230   240   250
GTCTCCTGAGCATTGCTCACCTCACCATACTGCACTCAGGCAAGCCATTC

MGR61

   260   270   280   290   300
TCTGCTGGGGGGAATTGATGACTCTAGCTACCTGGGTGGGTAATAATTTG

   310   320   330   340   350
GAAGATCCAGCTAGCAGGGATCTAGTAGTCAATTATGTTAATACTAACAT

   360   370   380   390   400
GGGTTTAAAGATCAGGCAACTATTGTGGTTTCATATATCTTGCCTTACTT

   410   420   430   440   450
TTGGAAGAGAGACTGTACTTGAATATTTGGTCTCTTTCGGAGTGTGGATT

   460   470   480   490   500
CGCACTCCTCCAGCCTATAGACCACCAAATGCCCCTATCTTATCAACACT

   510   520   530   540   550
TCCGGAAACTACTGTTGTTAGACGACGGGGCCGAGGCAGGTCCCCTAGAA
MGR371 ACTAGT   MGR62
       Spe1   MGR369   ACTAGT
                              Spe1

   560   570   580   590   600
GAAGAACTCCCTCGCCTCGCAGACGCAGATCTCAATCGCCGCGTCGCAGA

   610   620   630
AGATCTCAATCTAGGGAATCTCAATGTTAG

ACTAGT   MGR370
Spe1

Fig.3A.

```
CTGTATTTCCCTGCTGGTGGCTCCAGTTCAGGAACAGTAAACCCTGTTCTGACTACTGCCTCTCCCTTATCGTCAATCTTCTCGAGGATTGGGGACCCTGCCGCTGAAC
 L  Y  F  P  A  G  G  S  S  S  G  T  V  N  P  V  L  T  T  A  S  P  L  S  S  I  F  S  R  I  G  D  P  A  L  N
```

Fig.3B.

```
ATTCCAGGAACAACAACAACCAGTACGGACCCATGCAAAAACCTGCACGACTCCTGCTCAAGGCAACTCTATGTTCCCTCATGTTGCTGTACAAAACCTTCGGATGGA
 I  P  G  T  T  T  T  S  T  G  P  C  K  T  C  T  T  P  A  Q  G  N  S  M  F  P  S  C  C  C  T  K  P  S  D  G
AATTGC
 N  C
```

Fig.3C.

```
ATTCCAGGAACAACAACAACCAGTACGGGACCCATGCAAAAACCTGCGACGACTCCTGCTCAAGGCAACTCTATGTTCCCTCATGTTGCTGTACAAAACCTTCGGATGGA
 I  P  G  T  T  T  T  S  T  G  P  C  K  T  C  T  T  P  A  Q  G  N  S  M  F  P  S  C  C  C  T  K  P  S  D  G
AATTGCACCTGTATTCCCATCCCATCGTCTTGGGCT
 N  C  T  C  I  P  P  S  S  W  A
```

Fig.4.

←5Kb

Fig.5.

Fig.6.

1   2   3   4   5   6   7   8
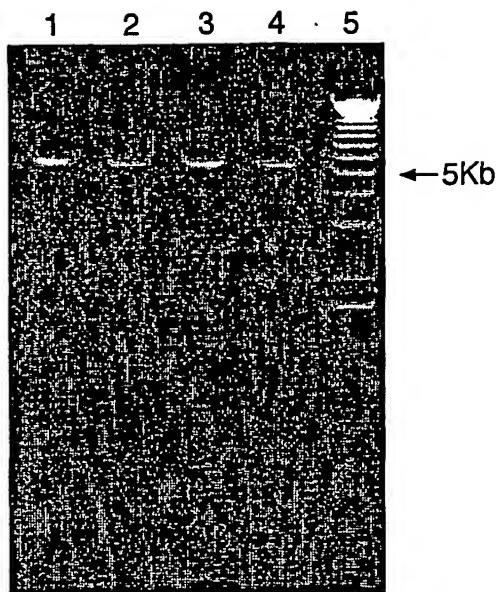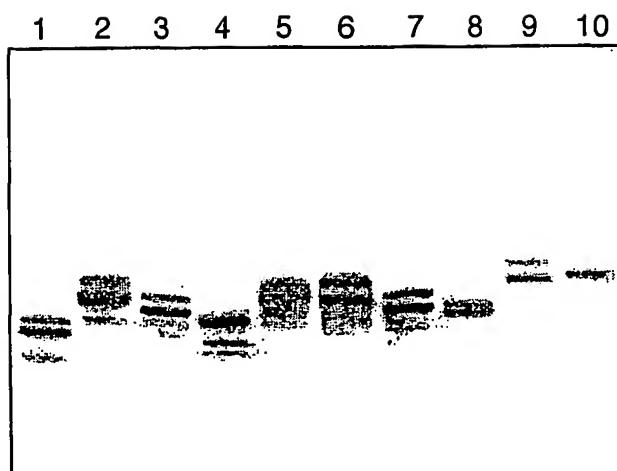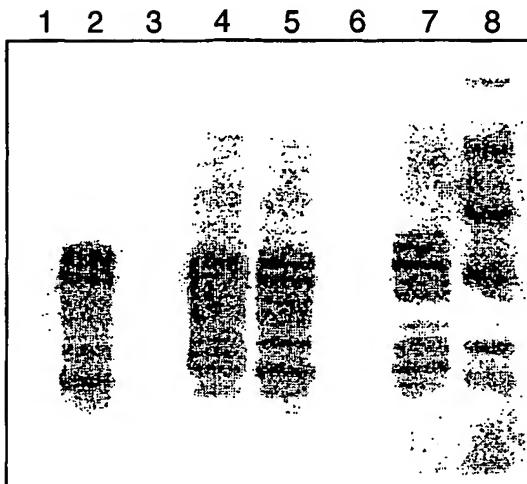
Fig.7.

1   2   3   4   5   6   7
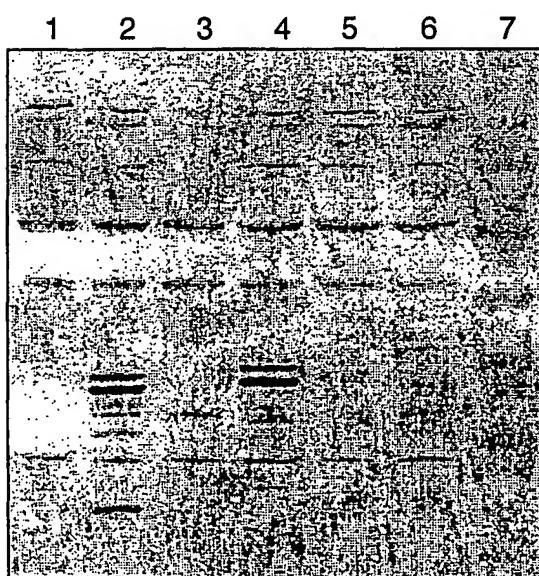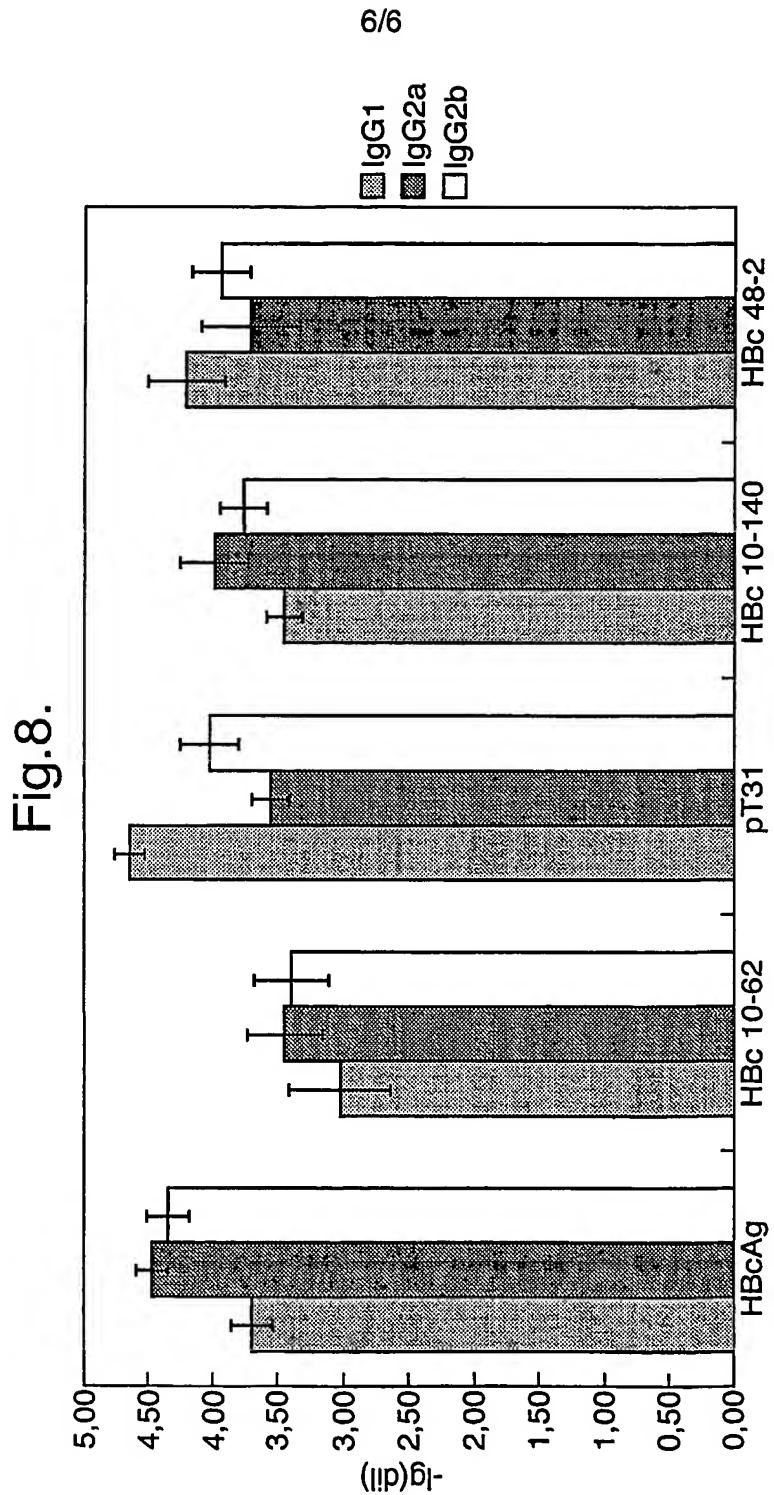
Fig.8.

SEQUENCE LISTING

<110> MEDEVA EUROPE LTD

<120> MODIFICATION OF HEPATITIS B CORE ANTIGEN

<130> P78451A

<160> 2

<170> PatentIn Ver. 2.1

<210> 1
<211> 639
<212> .DNA
<213> Hepatitis B virus

<220>
<221> CDS
<222> (1)..(639)

<400> 1
```
atg caa ctt ttt cac ctc tgc cta atc atc tct tgt tca tgt cct act    48
Met Gln Leu Phe His Leu Cys Leu Ile Ile Ser Cys Ser Cys Pro Thr
 1               5               10              15

gtt caa gcc tcc aag ctg tgc ctt ggg tgg ctt tgg ggc atg gac atc    96
Val Gln Ala Ser Lys Leu Cys Leu Gly Trp Leu Trp Gly Met Asp Ile
            20              25              30

gac cct tat aaa gaa ttt gga gct act gtg gag tta ctc tcg ttt ttg   144
Asp Pro Tyr Lys Glu Phe Gly Ala Thr Val Glu Leu Leu Ser Phe Leu
        35              40              45

cct tct gac ttc ttt cct tca gta cga gat ctt cta gat acc gcc tca   192
Pro Ser Asp Phe Phe Pro Ser Val Arg Asp Leu Leu Asp Thr Ala Ser
    50              55              60

gct ctg tat cgg gaa gcc tta gag tct cct gag cat tgt tca cct cac   240
Ala Leu Tyr Arg Glu Ala Leu Glu Ser Pro Glu His Cys Ser Pro His
65              70              75              80

cat act gca ctc agg caa gca att ctt tgc tgg ggg gaa cta atg act   288
His Thr Ala Leu Arg Gln Ala Ile Leu Cys Trp Gly Glu Leu Met Thr
```

```
                    85                    90                    95

cta gct acc tgg gtg ggt gtt aat ttg gaa gat cca gcg tct aga gac    336
Leu Ala Thr Trp Val Gly Val Asn Leu Glu Asp Pro Ala Ser Arg Asp
            100                 105                 110

cta gta gtc agt tat gtc aac act aat atg ggc cta aag ttc agg caa    384
Leu Val Val Ser Tyr Val Asn Thr Asn Met Gly Leu Lys Phe Arg Gln
            115                 120                 125

ctc ttg tgg ttt cac att tct tgt ctc act ttt gga aga gaa aca gtt    432
Leu Leu Trp Phe His Ile Ser Cys Leu Thr Phe Gly Arg Glu Thr Val
            130                 135                 140

ata gag tat ttg gtg tct ttc gga gtg tgg att cgc act cct cca gct    480
Ile Glu Tyr Leu Val Ser Phe Gly Val Trp Ile Arg Thr Pro Pro Ala
145                 150                 155                 160

tat aga cca cca aat gcc cct atc cta tca aca ctt ccg gag act act    528
Tyr Arg Pro Pro Asn Ala Pro Ile Leu Ser Thr Leu Pro Glu Thr Thr
                165                 170                 175

gtt gtt aga cga cga ggc agg tcc cct aga aga aga act ccc tcg cct    576
Val Val Arg Arg Arg Gly Arg Ser Pro Arg Arg Arg Thr Pro Ser Pro
            180                 185                 190

cgc aga cga agg tct caa tcg ccg cgt cgc aga aga tct caa tct cgg    624
Arg Arg Arg Arg Ser Gln Ser Pro Arg Arg Arg Arg Ser Gln Ser Arg
            195                 200                 205

gaa tct caa tgt tag                                                639
Glu Ser Gln Cys
        210


<210> 2
<211> 212
<212> PRT
<213> Hepatitis B virus

<400> 2
Met Gln Leu Phe His Leu Cys Leu Ile Ile Ser Cys Ser Cys Pro Thr
    1               5                   10                  15

Val Gln Ala Ser Lys Leu Cys Leu Gly Trp Leu Trp Gly Met Asp Ile
            20                  25                  30
```

Asp Pro Tyr Lys Glu Phe Gly Ala Thr Val Glu Leu Leu Ser Phe Leu
        35   40   45

Pro Ser Asp Phe Phe Pro Ser Val Arg Asp Leu Leu Asp Thr Ala Ser
        50   55   60

Ala Leu Tyr Arg Glu Ala Leu Glu Ser Pro Glu His Cys Ser Pro His
65   70   75   80

His Thr Ala Leu Arg Gln Ala Ile Leu Cys Trp Gly Glu Leu Met Thr
              85   90   95

Leu Ala Thr Trp Val Gly Val Asn Leu Glu Asp Pro Ala Ser Arg Asp
            100   105   110

Leu Val Val Ser Tyr Val Asn Thr Asn Met Gly Leu Lys Phe Arg Gln
        115   120   125

Leu Leu Trp Phe His Ile Ser Cys Leu Thr Phe Gly Arg Glu Thr Val
        130   135   140

Ile Glu Tyr Leu Val Ser Phe Gly Val Trp Ile Arg Thr Pro Pro Ala
145   150   155   160

Tyr Arg Pro Pro Asn Ala Pro Ile Leu Ser Thr Leu Pro Glu Thr Thr
            165   170   175

Val Val Arg Arg Arg Gly Arg Ser Pro Arg Arg Arg Thr Pro Ser Pro
        180   185   190

Arg Arg Arg Arg Ser Gln Ser Pro Arg Arg Arg Arg Ser Gln Ser Arg
        195   200   205

Glu Ser Gln Cys
        210

(51) International Patent Classification[7]: C12N 15/51, C07K 14/02, A61K 39/29, A61P 31/20

(21) International Application Number: PCT/GB01/02817

(22) International Filing Date: 22 June 2001 (22.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0015308.0   22 June 2000 (22.06.2000)   GB
0024544.9   6 October 2000 (06.10.2000)   GB

(71) Applicant (for all designated States except US): CELL-TECH PHARMACEUTICALS LIMITED [GB/GB]; 208 Bath Road Slough, Berkshire SL1 3WE (GB).

(72) Inventors; and
(75) Inventors/Applicants (for US only): PAGE, Mark [GB/GB]; 34 The Swallows, Welwyn Garden City, Hertfordshire AL1 1BY (GB). LI, Jing-Li [GB/GB]; 166 Ravenscroft Road, Beckenham, Kent BR3 4TW (GB). PUMPENS, Paul [LV/LV]; Biomedical Research and Study Centre, University of Lativa, Ratsupites Str. 1, LV-1067 (LV). BORISOVA, Galina [LV/LV]; Biomedical Research and Study Centre, University of Latvia, Ratsupites Str.1, Riga, Latvia 1067 (LV).

(74) Agents: CAMPBELL, Patrick, John, Henry et al.; J. A. Kemp & Co., 14 South Square, Gray's Inn, London WC1R 5JJ (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— with international search report

(88) Date of publication of the international search report:
28 March 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/98333 A3

(54) Title: MODIFICATION OF HEPATITIS B CORE ANTIGEN

(57) Abstract: A protein is provided comprising hepatitis B core antigen (HBcAg) wherein one or more of the four arginine repeats has been deleted, said protein comprising the C-terminal cysteine of HBcAg. The deleted region may be replaced by an epitope from a protein other than HBcAg, in which case the HBcAg acts as a carrier to present the epitope to the immune system. The chimeric protein is useful in prophylactic and therapeutic vaccination of a host, for example against hepatitis B virus.

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7   C12N15/51   C07K14/02   A61K39/29   A61P31/20

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7   C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, WPI Data, PAJ, MEDLINE, CHEM ABS Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HATTON T ET AL: "RNA AND DNA-BINDING ACTIVITIES IN HEPATITIS B VIRUS CAPSID PROTEIN A MODEL FOR THEIR ROLES IN VIRAL REPLICATION" JOURNAL OF VIROLOGY, vol. 66, no. 9, 1992, pages 5232-5241, XP001041852 ISSN: 0022-538X page 5232, left-hand column, paragraph 2 -right-hand column, paragraph 1 page 5232, right-hand column, paragraph 3 -page 5233, left-hand column, paragraph 2 page 5234; figure 1 page 5234, left-hand column, paragraph 1 -right-hand column, paragraph 1 page 5235, right-hand column, paragraph 3 page 5238; figure 7 --- -/--	1,3-5, 13,17-22

[X] Further documents are listed in the continuation of box C.   [X] Patent family members are listed in annex.

* Special categories of cited documents :

'A' document defining the general state of the art which is not considered to be of particular relevance

'E' earlier document but published on or after the international filing date

'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

'O' document referring to an oral disclosure, use, exhibition or other means

'P' document published prior to the international filing date but later than the priority date claimed

'T' later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

'X' document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

'Y' document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

'&' document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
7 December 2001	21/12/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Sitch, W

Form PCT/ISA/210 (second sheet) (July 1992)

2

In' ·ational Application No

PCT/GB 01/02817

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	WO 00 26385 A (POWDERJECT VACCINES INC) 11 May 2000 (2000-05-11) page 2, paragraph 1 page 26 -page 28; example 1 page 28 -page 30; example 2 ---	2
A	SCHOEDEL F ET AL: "HYBRID HEPATITIS B VIRUS CORE ANTIGEN AS A VACCINE CARRIER MOIETY: I. PRESENTATION OF FOREIGN EPITOPES" JOURNAL OF BIOTECHNOLOGY, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 44, no. 1/3, January 1996 (1996-01), pages 91-96, XP000891474 ISSN: 0168-1656 page 91 abstract ---	
A	EP 0 421 635 A (WELLCOME FOUND) 10 April 1991 (1991-04-10) cited in the application page 2, line 35 - line 46 ---	
A	BORISOVA G P ET AL: "RECOMBINANT CORE PARTICLES OF HEPATITIS B VIRUS EXPOSING FOREIGN ANTIGENIC DETERMINANTS ON THEIR SURFACE" FEBS LETTERS, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 259, no. 1, 1 December 1989 (1989-12-01), pages 121-124, XP002035980 ISSN: 0014-5793 page 122, left-hand column, paragraph 5 -page 124, left-hand column, paragraph 1 ---	
A	HUI ERIC KA-WAI ET AL: "Hepatitis B virus maturation is affected by the incorporation of core proteins having a C-terminal substitution of arginine or lysine stretches." JOURNAL OF GENERAL VIROLOGY, vol. 80, no. 10, October 1999 (1999-10), pages 2661-2671, XP002185072 ISSN: 0022-1317 page 2661 abstract page 2663; figure 1 ---	

-/--

2

	Int'l Application No
	PCT/GB 01/02817

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	DATABASE BIOSIS 'Online! BIOSCIENCES INFORMATION SERVICE, PHILADELPHIA, PA, US; 1992 ZHOU S ET AL: "CYS RESIDUES OF THE HEPATITIS B VIRUS CAPSID PROTEIN ARE NOT ESSENTIAL FOR THE ASSEMBLY OF VIRAL CORE PARTICLES BUT CAN INFLUENCE THEIR STABILITY" Database accession no. PREV199294115400 XP002185073 abstract & JOURNAL OF VIROLOGY, vol. 66, no. 9, 1992, pages 5393-5398, ISSN: 0022-538X -----	

2

INTERNATIONAL SEARCH REPORT

Information on patent family members

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 0026385	A	11-05-2000	AU	2022200 A	22-05-2000
			EP	1119630 A1	01-08-2001
			WO	0026385 A1	11-05-2000
EP 0421635	A	10-04-1991	AU	626183 B2	23-07-1992
			AU	6269090 A	11-04-1991
			CA	2025598 A1	20-03-1991
			DD	298134 A5	06-02-1992
			DE	69021002 D1	24-08-1995
			DE	69021002 T2	23-11-1995
			DK	421635 T3	27-11-1995
			EP	0421635 A1	10-04-1991
			ES	2075883 T3	16-10-1995
			IE	903370 A1	10-04-1991
			JP	3216186 A	24-09-1991
			NZ	235315 A	25-09-1991
			ZA	9007233 A	27-05-1992

Intervirology

HBV Core Particles as a Carrier for B Cell/T Cell Epitopes

Paul Pumpens   Elmars Grens

Biomedical Research and Study Center, University of Latvia, Riga, Latvia

Key Words

Virus-like particles · Hepatitis B virus core particles · Chimeric proteins · Self-assembly · Molecular display · Epitopes · Antigenicity · Immunogenicity

Abstract

In the middle 80s, recombinant hepatitis B virus cores (HBc) gave onset to icosahedral virus-like particles (VLPs) as a basic class of non-infectious carriers of foreign immunological epitopes. The recombinant HBc particles were used to display immunodominant epitopes of hepatitis B, C, and E virus, human rhinovirus, papillomavirus, hantavirus, and influenza virus, human and simian immunodeficiency virus, bovine and feline leukemia virus, foot-and-mouth disease virus, murine cytomegalovirus and poliovirus, and other virus proteins, as well as of some bacterial and protozoan protein epitopes. Practical applicability of the HBc particles as carriers was enabled by their ability to high level synthesis and correct self-assembly in heterologous expression systems. The interest in the HBc VLPs was reinforced by the resolution of their fine structure by electron cryomicroscopy and X-ray crystallography, which revealed an unusual α-helical organization of dimeric units of HBc shells, alternative packing into icosahedrons with T = 3 and T = 4 symmetry, and the existence of long protruding spikes. The tips of the latter seem to be the optimal targets for the display of foreign sequences up to 238 amino acid residues in length. Combination of numerous experimental data on epitope display with the precise structural information enables a knowledge-based design of diagnostic, and vaccine and gene therapy tools on the basis of the HBc particles.

Introduction

Hepatitis B core (HBc) particles were first reported as a promising virus-like particle (VLP) carrier in 1986 [1] and published in 1987 [2, 3]. Being one the first VLP candidates and the first icosahedral VLP carrier, the HBc particles remain the most flexible and the most promising model for knowledge-based display of foreign peptide sequences up to now. The use of HBc particles as a VLP carrier has been reviewed extensively. For detailed analyses, we recommend specialized reviews dealing with the role of HBc particles as components of HBV infection [4–6] and as a VLP carriers [7–12].

In many ways, HBc protein holds a unique position among other VLP carriers because of its high-level expression and efficient particle formation in virtually all known homologous and heterologous expression systems, includ-

ing bacteria. Correct folding of the HBc protein and formation of authentic HBc particles have been documented in various mammalian cell cultures [13–18], retrovirus [19], vaccinia virus [20, 21] and adenovirus [22] expression systems, frog *Xenopus* oocytes [23], insect *Spodoptera* cells [24–27], yeast *Saccharomyces cerevisiae* [28–31], in plants *Nicotiana tabacum* [32], and in bacteria such as *Escherichia coli* [33–42], *Bacillus subtilis* [43], *Salmonella* [44] and *Acetobacter* [45]. Electron microscopy revealed the ultrastructural identity of the HBc particles derived from either HBV virions and infected hepatocytes, or from *E. coli* [46] or yeast [47]. Moreover, comparative electron cryomicroscopy and three-dimensional image reconstruction of HBV cores of natural and bacterial origin reconfirmed the native HBc structure in bacteria at the molecular level, in the absence of the complete viral genome and other viral components [48].

Intrinsic Properties of the HBc Particle

Biological Multifunctionality of the HBc Protein

The natural multifunctionality of the HBc protein seems to be responsible for its unusual flexibility, which is advantageous for its usage as a VLP carrier. Although the HBV gene C has only two in-frame initiation AUG codons (fig. 1a), it is responsible for the appearance of at least four different polypeptides: p25, p22, p21, and p17 [for a review, see ref. 49]. The p25 precore protein, starting at the first AUG codon, becomes targeted by a signal peptide in the preC sequence to a cell secretory pathway, in which a p22 is formed by N-terminal processing. The p22 undergoes further cleavage at the C-terminal region, after position 149, to generate a p17 protein, or HBe protein, which is secreted from the cell as the HBe antigen.

The predominant p21 polypeptide is synthesized from the second AUG of the open reading frame, and constitutes a structural component of the HBcAg, or HBV nucleocapsid, and may therefore be referred to as a genuine HBc polypeptide. The HBc polypeptide is able to self-assemble and was therefore selected as a target for protein engineering manipulations. Besides capsid-building, the p21 protein participates in the viral life cycle and its regulation, including the synthesis of double-stranded DNA as a cofactor of the viral reverse transcriptase-DNA polymerase, viral maturation, recognition of viral envelope proteins and budding from the cell [for details see ref.

which inhibits the binding of the L protein to the HBc particle is located at the tips of the spikes of the latter [53].

Phosphorylation of serine residues by cellular protein kinase C within three repeated SPRRR motifs on the C-terminus of the p21 protein [54–56] and its role in the maturation of the HBV capsid [57, 58] reflect the complex function of the HBc particles. According to recent data, phosphorylation of HBc subunits induces a conformational change that exposes the C-terminal sequences, which may protrude through the holes in the capsid wall and become accessible on the surface to serve as a nuclear targeting signal [58].

Fine Structure of the HBc Particle

In contrast to HBsAg, representing a complex and irregular lipoprotein structure, HBcAg consists of 180 or 240 copies of identical polypeptide subunits.

The fine structure of HBc particles (fig. 1b) was revealed by electron cryomicroscopy and image reconstruction [59–61] and finally resolved by X-ray crystallography at 3.3 Å resolution [62]. The organization of HBc particles was found to be largely α-helical and quite different from previously known viral capsid proteins with β-sheet jelly-roll packings [59, 62]. Association of two amphipathic α-helical hairpins results in the formation of a dimer with a four-helix bundle as the major central feature (fig. 1c). The dimers are able to assemble into two types of particles, large and small ones, which are 34 and 30 nm in diameter and correspond to triangulation number T = 4 and T = 3 packings, containing 240 and 180 HBc molecules, respectively. The four-helix bundles protrude, forming spikes approximately 25 Å in length and 20 Å in width [62]. The amino acid stretch 76–81 located at the tips of the spikes presents a central part of the so-called major immunodominant region (MIR) of the HBc particle.

In addition to the MIR, the region 127–133 is the next exposed and accessible epitope on the particle surface. This region is located at the end of the C-terminal α-helix and forms small protrusions on the surface of the HBc particle.

Although the C gene is the most conserved amongst HBV genes, numerous amino acid substitutions were fixed for its most parts. A portrait of the C protein, with elements of its three-dimensional structure and distribution of mutations, is given in figure 2. It is evident that only the N-terminal region 1–11 and some special posi-
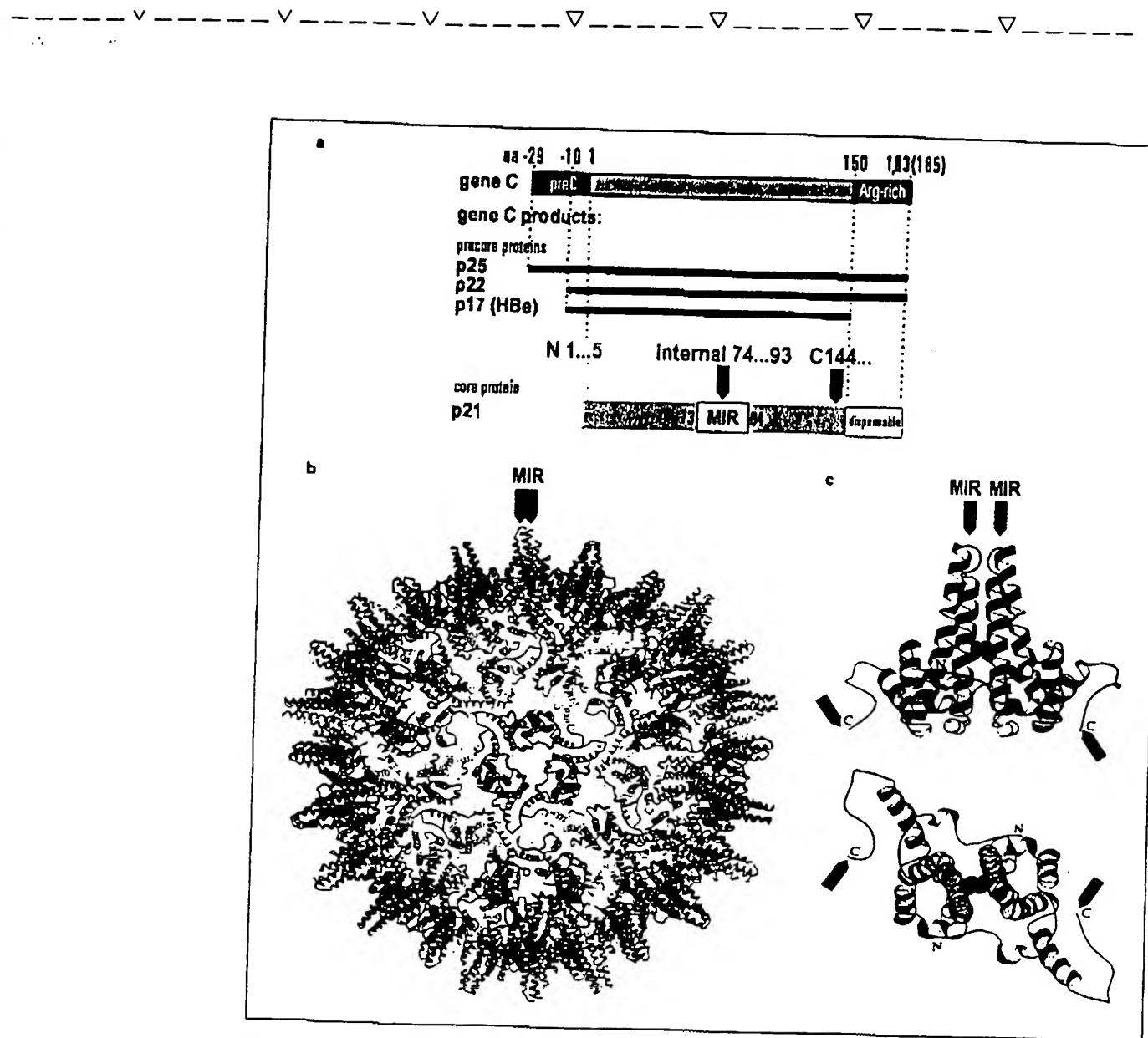
Fig. 1. HBc particles as carriers for foreign epitopes. **a** Products encoded by the C gene with localization of the insertion sites for the foreign epitopes. **b** Two orthogonal views of the HBc dimer (subunits C and D) viewed normal to the local two-fold axis and along the two-fold axis from the outside of the capsid [62]. Cys-61, which forms a disulfide bridge between the two monomers, is shown in green, and Cys-48, which does not form disulfides, in yellow. Insertion sites for foreign epitopes are marked by arrows. **c** T = 4 HBc capsid viewed down an icosahedral three-fold axis [62]. The HBc maps are a generous gift of R.A. Crowther.

acid positions of the α1, α2a/α2b, α5 (proximal part) helices and especially of the MIR (including the α4a and the proximal part of the α4b) are able to accept amino acid changes.

Of particular structural value was the clear demonstration of dispensability of the C-terminal protamine-like arginine-rich domain of the p21 protein (aa 150–183) for its self-assembly capabilities in the so-called HBcΔ particles [63–65]. The HBcΔ particles formed by C-terminally
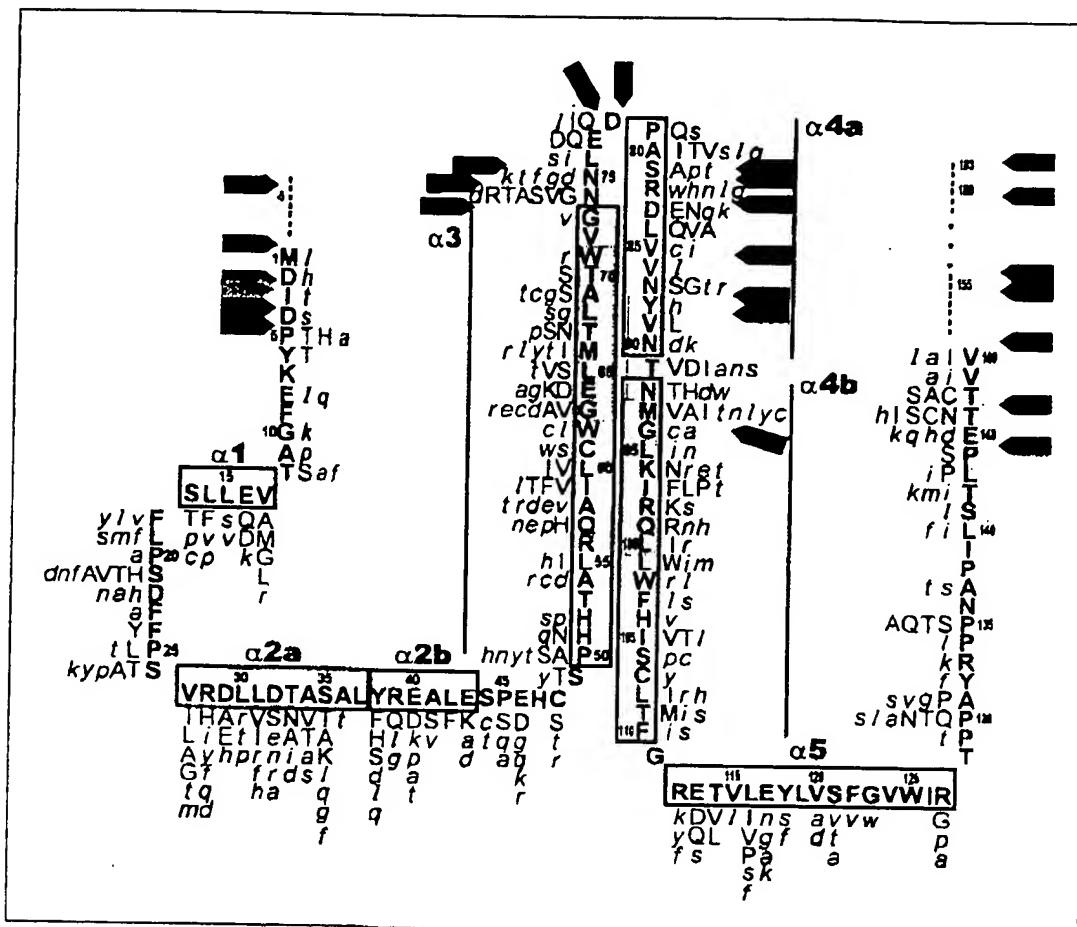
Fig. 2. Diversity of the primary structure of the HBc molecule. The amino acid sequence of the CW variant of HBc particle resolved by X-ray crystallography [62] is given as a default sequence. α-Helices derived from the HBc crystal structure [62] are boxed. Amino acid substitutions from more than 200 HBV structures available in sequence data bases (GenBank, SwissProt) and from more than 100 HBV structures presented in original publications are compiled. Unique amino acid substitutions found in no more than one HBc sequence are indicated in lower case italics. Amino acid substitutions suspected by authors in connection with the particular features in the course of HBV infection are shown in red. B cell and CTL epitopes are colored yellow and pink, respectively. Blue arrows show the putative insertion sites and boundaries for 'permitted' deletions/substitutions. Pink arrow locates the natural N-terminal insertion of dodeca-peptide RTTLPYGLPGLD within the C gene of the HBV genotype G [180].

truncated polypeptides were practically indistinguishable from the HBc particles formed by full-length HBc polypeptides, as shown by electron cryomicroscopy [59]. However, HBcΔ particles were less stable, failed to encapsidate nucleic acid and accumulated usually as empty shells, in contrast to the full-length HBc particles [59, 63, 66–68]. The unusual molecular flexibility of the C-terminal protamine-like domain has been revealed by the attempts to apply NMR spectroscopy to structural analy-

sis of HBc particles [68]. The C-terminal limit for self-assembly of HBcΔ particles was mapped experimentally between aa residues 139 and 144 [65, 67, 69]. According to more recent data [70], this border maps at position 140, and the appropriate HBcΔ particles form predominantly the T = 3 isomorph with a proportion of T = 4 isomorph of approximately 18%. The proportion of T = 4 capsids increases with the length of HBc polypeptide, and the HBc variants truncated at positions 142, 147, and 149 aa

form about 52. 79 and 94% of T = 4 capsids. respectively. In this respect, HBc particles remain the behavior of yeast Ty particles. The length of the C-terminal region of the Ty monomer was found to dictate the T number, and thus the size. of the assembled particles in the broad range from T = 3 to T = 9 shells [71].

The HBcΔ particles played the most important role in three-dimensional resolution of the HBc shells [62] and seem to be the most promising candidates for further vaccine and gene therapy applications.

Intrinsic Epitopes of the HBc Particle

The extremely high immunogenicity of HBc particles has been known for a long time. in contrast to the relatively low immunogenicity of HBV envelope proteins. Thus, HBV patients develop a strong and long-lasting humoral anti-HBc response [72]. Among the HBV polypeptides, HBc induces the strongest B-cell, T-cell and cytotoxic T lymphocyte (CTL) response [for a review, see ref. 73]. HBc particles are known to function as both T-cell-dependent and T-cell-independent antigens [74]. Following immunization. it primes preferentially Th1 cells. does not require an adjuvant [75, 76]. and is able to mediate an anti-HBs response [77]. Recently. the enhanced immunogenicity of HBc particles was explained by their ability to be presented by B cells as a primary antigen to T cells in mice [78]. HBc particles elicit a strong CTL response during HBV infection [79], and this response is maintained for decades following clinical recovery, apparently keeping the virus under control [80].

The major B-cell epitopes c (HBc epitope) and e1 (HBe epitope 1) are localized within the MIR of the HBc protein. around the protruding region 76–81. lying on the tip of the spike [81, 82]. and cover the loop α3/α4a and the α4a helix. The next important epitope e2 (HBe epitope 2) lies on the other surface-exposed region of the HBc protein adjacent to the α5 helix. around positions 129–132 [83, 84].

HLA-class-II-restricted. T helper cell epitopes of the HBc protein are revealed to peptides 1–20. 28–47, 50–69. 72–90. 81–105. 90–99. 108–122. 111–125. 117–131, 120–139. 126–146. and 141–165 [73, 85–87]. In mice. the following sequences were documented among the T cell epitopes: 120–140 (haplotype H-2s.b). 100–120 (haplotype H-2f.q). and 85–100 (H-2d mice) [88]. The sequence 120–140 was further subdivided into two significant parts 120–131 and 129–140 stimulating B10.S (H-2s) and B10 (H-2b) HBc-primed T cells. respectively [89].

Since recent studies in HBV-infected patients have suggested that hepatocytolysis induced by CD8+ CTLs is the most important effector pathway in eliminating infected cells. special attention was devoted to search for HLA class I-restricted CTL epitopes within the HBc molecule. In men, practically a single HLA-A2-restricted epitope 18–27 has been identified, containing the predicted HLA-A2 binding motif with Leu at position 2 and Val at the C-terminus [90]. An HBc epitope 141–151 has been defined by CTL clones from patients with acute hepatitis B. that is restricted by both HLA-Aw68 and HLA-A31 molecules [91]. Peptide 88–96. sharing the HLA-A11 binding motifs and recognized by HLA-A11-restricted CD8+ CTLs, was isolated directly from HLA class I molecules of HBV-infected liver cell membrane [92].

In mice, the HBc peptides 93–100 [93] and 87–96 [94] were found as CTL epitopes in the context of Kb-binding (H-2b mice) and Kd-binding (H-2d mice). respectively. In macaques. the long-lived CTL response was directed against HBc peptide 63–71 [95].

The location of HBc epitopes is shown in figure 2. except of T cell epitopes. which cover the HBc molecule practically at full length.

Display of Foreign Epitopes on the HBc Particle

In general. it is widely accepted now that the HBc carrier is capable of ensuring a high level of B cell and T cell immunogenicity to foreign epitopes [7–12]. In addition to the ability of the HBc carrier moiety to provide T cell help to inserted sequences. the HBc capsid mediates the T-cell-independent character of the humoral response to inserted epitopes. due to the high degree of repetitiveness of the epitopes and the proper spacing between them [96].

Experimental search for the appropriate target sites for foreign insertions pointed to the N- and C-termini of the HBc molecule, as well as to its MIR at the tip of the spike [7–12]. These findings are in a good agreement with the X-ray data (fig. 1). because these regions do not participate in the critical intra- and intermolecular interactions [62]. General characteristics of chimeric HBc derivatives are compiled in table 1.

N-Terminal Insertions

Historically. N-terminal insertions were the first ones. in which chimeric HBc particles carrying the VP1 epitope 141–160 of foot-and-mouth disease virus (FMDV) were demonstrated in vaccinia virus expression system [1, 2]. and yeast [97]. The ability of the HBc chimera to induce FMDV-neutralizing antibodies stimulated authors to construct other N-terminal insertion variants with epi-

topes of the gp70 protein of feline leukemia virus. VP2 protein of human rhinovirus type 2, VP1 protein of poliovirus type 1 [98, 99]. Env protein of simian immunodeficiency virus [100]. outer membrane protein P.69 (pertactin) from bacteria *Bordetella pertussis* [101]. and chorionic gonadotropin [97]. The latter was constructed as a contraceptive vaccine candidate.

The N-terminus of HBc molecule was used also as a target for insertion of relatively short epitopes from HBV preS [102–104]: HIV-1 gp120 and p24 [105]. gp41, p34 Pol. and p17 Gag [106, 107]: and from human cytomegalovirus gp58 [108]. The latter could not be purified or characterized immunologically, although it formed VLPs.

Fusion of 45 N-terminal aa of the Puumala hantavirus nucleocapsid protein to the N-terminus of HBcΔ allowed the formation of chimeric VLPs, which induced a strong antibody response and some protection in the bank vole model [109]. However, addition of 120 N-terminal aa of the hantavirus nucleocapsid to the N-terminus of HBcΔ prevented self-assembly, in contrast to their insertion into position 78 [110] (see below).

The recent remarkable breakthrough in the application of the HBc model for vaccine development was based on the N-terminal insertion. Chimeric particles expressed in *E. coli* and carrying 23 aa of the extracellular domain of influenza A minor protein M2 (the initiating methionine was completely removed after expression) provided up to 100% protection against a lethal virus challenge in mice, after intraperitoneal or intranasal administration [111]. This protection was mediated by antibodies.

In general. N-terminal insertions seemed to be displayed on the surface of the HBc particle [112] and assured a high level of antibody response to inserted epitopes. Deletions of more than 4 aa residues at the N-terminus of the HBc molecule result in a protein, which is not competent for self-assembly. The capacity of N-terminal HBc vectors is around 50 aa, the inserted epitopes are accessible to specific antibodies.

Internal Insertions

The MIR, or tip of the spike of the HBc molecule, is generally accepted now as a target site of choice. Insertion of foreign epitopes into the MIR guarantees a high level of specific B cell and T cell immunogenicity. In spite of its internal location, the MIR allows for a surprisingly high capacity of insertions. For example, the entire 120 aa long immunoprotective region of the hantavirus nucleocapsid was inserted into the MIR of the C-terminally truncated HBcΔ particles, whereas N- and C-termini failed to accept

this fragment for self-assembly [110]. It is necessary to emphasize that the shorter, aa 1–45 segment of hantavirus nucleocapsid within the MIR also ensured protection of bank voles against virus challenge after immunization with chimeric particles [113, 114].

Moreover. green fluorescent protein of 238 aa was natively displayed on the surface of full-length HBc particles [115]. Chimeras demonstrated not only fluorescence capabilities. but also elicited a potent humoral response against native GFP. This example shows the structural importance of proper and independent folding of sequences subjected to exposure on the HBc particles and opens the way for high-resolution structural analyses of nonassembling proteins by electron microscopy [115].

Historically, the story of MIR insertions started with the introduction of up to 27 aa long epitopes of HBV preS [104, 116–120]. 18 aa of VP2 protein from the human rhinovirus type 2 [99, 121]. up to 30 aa of the simian immunodeficiency virus Env [100]. and 25 aa [122, 123] and up to 43 aa [119, 120] of the V3 loop of the HIV-1 gp120. Insertion of 39 aa of the domain 'a' sequence from the HBsAg (positions 111–149) was the first successful attempt to mimic a conformational epitope on the surface of chimeric particles [124].

HBsAg and preS epitopes have been chosen for the construction of first multivalent particles. namely for simultaneous insertion of different foreign sequences from the preS1 and preS2 regions into the MIR and into the N-terminus [125]. or into the MIR and into the C-terminus [104, 117]. or from the HBsAg and preS2 into the MIR and into the C-terminus of the HBc protein [119]. respectively. Later. multivalent HBcΔ particles carrying different hantavirus nucleocapsid epitopes at the MIR and C-terminus were constructed [114].

First mosaic HBc particles carrying chimeric and wild-type HBc monomers were also constructed on the basis of full-length HBc vector for internal insertions. In this case, an epitope of 8 aa from the Venezuelan equine encephalomyelitis virus E2 protein has been inserted into position 81 of the HBc molecule [126].

An attempt to construct a therapeutic vaccine against HPV16-associated anogenital cancer was undertaken by MIR insertions of B cell, T cell, and CTL epitopes from the E7 oncoprotein of the human papillomavirus type 16 [127–129]. Humoral and T-proliferative responses to the chimeras were elicited successfully [127]. also in the case of *Salmonella*-driven expression ([128]. see below). but the appropriate chimeric particles failed to prime E7-directed CTL responses in mice [129].

Table 1. HBc particles as VLP carriers of foreign epitopes

Insertion site	Source and properties of the insertion				Expression system	Major immunological activity	References
	species	protein	epitope	length of insertion			
N-terminal insertions							
Full-length HBc carrier −6	FMDV	VP1	143–160	36	S. cerevisiae vaccinia	B (guinea pigs)	1,2,97
	FeLV	gp70	137–153	31	E. coli	B (guinea pigs)	98
	HIV-1	gp120	303–327	52	E. coli	–	105
		p24	288–304	24	E. coli	–	105
	SIV	gp120	170–189	31	E. coli	B (guinea pigs)	100
		TMP	655–675	32	E. coli	B (guinea pigs)	100
	HRV-2	VP2	156–170	30	E. coli	B (guinea pigs)	98,99
	PV1	VP1	93–103	24	E. coli	B (guinea pigs)	98,99
	H. sapiens	hCG	109–145	50	S. cerevisiae	B (guinea pigs)	97
	B. pertussis	P.69	571–600	31	S. cerevisiae	B (mice)	101
−4	HBV	preS1	12–47	41	E. coli	B (mice)	104
−1	HBV	preS1	27–53	28	E. coli	B (mice)	104
5	Influenza A	M2	1–24	24	E. coli	B (mice)	110
C-terminally truncated HBcΔ carrier 1	Hanta	N	1–45	57	E. coli	–	109
2	HBV	preS1	31–35	22	E. coli	–	103
3	HBV	preS1	31–36	13	E. coli	B (rabbits, mice)	102,125
			94–105	19	E. coli	B (rabbits, mice)	102,125
			3×(94–105)	40	E. coli	B (rabbits, mice)	102,125
		preS2	133–143	14,18	E. coli	B (rabbits, mice)	102,125
	HIV-1	gp41	593–604	23	E. coli	B (rabbits)	106
		p34	940–949	16,21	E. coli	B (rabbits)	106
		p17	99–115	27	E. coli	B (rabbits)	106
Internal insertions							
Full-length HBc carrier 81/82	HRV2	VP2	156–170	18	E. coli	B (guinea pigs)	99,121
	SIV	gp120	121–147	30	E. coli	B (guinea pigs)	100
		TMP	655–675	23	E. coli	B (guinea pigs)	100
		TMP	738–763	26	E. coli	B (guinea pigs)	100
	FMDV	VP1	135–160	28	E. coli	B (guinea pigs)	134
	T. annulata	SPAG-1	785–892	110	E. coli	B, T (calves)	131,132
	HPV	E7	10–14	7	E. coli	B (mice)	127
			35–54	22	E. coli	B, T (mice)	127
			10–14 + 35–54	26	E. coli	–	127
			10–14 + 82–90 + 86–93	34	E. coli	B, no CTL (mice)	129
			10–14 + 86–93	15	S. typhimurium	B, no CTL (mice)	129
	VEE	E2	233–240	8	E. coli	–	126
	HBV	HBsAg	137–147	11	S. typhimurium	B, T (mice)	178,179
82/83	HBV	preS2	133–143	11	E. coli	–	F.S.116
C-terminally truncated HBcΔ carrier 73–82	HBV	preS1	31–35	11	E. coli	–	135,136
75–81	HBV	preS1	27–53	27	E. coli	B (mice)	135,136
	P. falciparum	CS	4×(NANP)	16	S. typhimurium	B, T (mice)	130
	P. berghei	CS	2×(DP₁NPN)₂		S. typhimurium	B, T (mice)	130
78/79	HBV	preS1	31–36	8	E. coli	B, T (mice)	135,136
			31–35	10	E. coli	–	103,135,136
	HIV-1	gp120	107–131	28	E. coli	–	119,120
	Hanta	N	1–45	55	E. coli	B (bank voles)	113,114
			1–120	130	E. coli	–	109
78–81	HIV-1	gp120	303–327	25	E. coli	B (mice)	122,123
			299–338	43	E. coli	B (mice)	122,123
			306–328	26	E. coli	B (mice)	122,123
	A. victoria	GFP	1–238	257	E. coli	B (rabbits)	115
78–82	HBV	preS1	31–35	12	E. coli	B, T (mice)	135,136
		HBsAg	111–149	41	E. coli	B, T (mice)	124
78–86	HBV	preS1	31–35	8,11	E. coli	B, T (mice)	135,136
78–89	HBV	preS1	31–35	7,13	E. coli	B, T (mice)	135,136
78–94	HBV	preS1	31–35	7	E. coli	B (mice)	135,136
81/82	HBV	preS1	31–36	26	E. coli	B (rabbits, mice)	125
			94–105	32	E. coli	B (rabbits, mice)	125

(handwritten marginal annotations: "Clarke", "CSH '90", "CSH '97", "deletion", "J04.117 CSH", "JEM", circled "81")

Table 1 (continued)

Insertion site		Source and properties of the insertion			length of insertion	Expression system	Major immunological activity	References
		species	protein	epitope				
C-terminal insertions								
Full-length HBc carrier	144	145	HBV						
				preS1	31–35	17	E. coli	–	
					31–80	61	E. coli	–	103
					80–118	46	E. coli	B. T (mice)	140–142
				preS2	118–173	62	E. coli	–	140–142
	179	180	HIV-1	gp41	589–640	60	E. coli	B. T (mice)	140–142
	183		MCMV	pp89	168–176	11	E. coli	–	140, 143, 144
			HBV	preS1	31–34	10	E. coli	CTL (mice)	145
			SIV	env	170–189	20	E. coli	–	103, 151
					324–339	16	E. coli	no B (guinea pigs)	100
					594–616	23	E. coli	no B (guinea pigs)	100
					655–675	21	E. coli	no B (guinea pigs)	100
C-terminally truncated HBcΔ carrier	144		BLV	gp51	89–137	50	E. coli	no B (guinea pigs)	100
			FMDV	VP1	200–213+	56	E. coli	–	140, 143, 150
			HBV	preS1	131–160		E. coli	B (rabbits)	153
					31–35	24	E. coli	–	103
					12–31	34	E. coli	B (rabbits)	146
					12–47	45	E. coli	B (rabbits)	146
					31–79	67	E. coli	B. T (mice)	140–142
				preS2	118–173	68	E. coli	B. T (mice)	140–142
					124–174	70	E. coli	B. T (mice)	140–142
					120–145	36	E. coli	B (rabbits)	146
				HBsAg	111–156	58	E. coli	B. T (rabbits)	12, 146, 147
					111–165	65	E. coli	B. T (rabbits)	12, 146, 147
				preS1	1–20+	122	E. coli		12
				preS2	1–26+				
				HBsAg	111–156				
			HBV + HCV	preS1	1–20+	144	E. coli	B (rabbits, mice)	156
				preS2	1–26+				
				core	1–98				
			HCV	core	6–77	72	E. coli	–	155
					6–143	138	E. coli	–	155
				NS3	1359–1449	91	E. coli	–	157
					1460–1532	73	E. coli	–	157
			HIV-1	gp120	299–338	46	E. coli	B (mice)	120
					306–328	29	E. coli	B (rabbits, mice)	120, 148
				gp41	616–632	20	E. coli	B (rabbits)	146
					667–680	16	E. coli	B (rabbits)	146
					728–751	26	E. coli	B (rabbits)	146
					589–640	60	E. coli	–	140
				gag	121–210	99	E. coli	B (mice)	148, 149
				nef	113–130	20	E. coli	–	143
			Hanta	N	1–45	62, 71	E. coli	–	109
					38–82	60	E. coli	B (bank voles)	113, 114
					75–119	60	E. coli	B (bank voles)	109, 113, 114
					1–114²	140	E. coli	B (bank voles)	114, 162
					1–120²	146	E. coli	–	110
			HCMV	gp58	599–644	46	E. coli	B (rabbits)	108
	146		HEV	ORF2	613–654	47	Baculovirus	–	133
	149		HCV	core	39–75	39	E. coli	–	154
					1–91	93	E. coli	–	154
					1–180	193	E. coli	–	154
					2×(1–180)	377	E. coli	–	154
					3×(1–180)	559	E. coli	–	154
					4×(1–180)	741	E. coli	–	154
	154		HIV-1	gp120	303–327	26	E. coli	B (mice)	122, 123
	155		S. aureus	nuclease	1–149	167	E. coli	–	159
	156		HBV		133–143	11	E. coli / S. typhimurium	B (mice)	104, 116
	157		P. gingivalis	Rgp-1	865–911	47	E. coli	B (rabbits, mice)	158

Only chimeras, which were found self-assembly competent, are included.

¹ This series is based on HBc vector truncated after aa position 176. ² Chimeras self-assemble only in the presence of wt HBc in the form of mosaic particles.

MIR insertions into the HBc particle were thoroughly investigated for construction of possible vaccines against infectious diseases caused by intracellular parasites. First, against malaria, in which chimeric HBcAg particles carrying circumsporozoite (HBcAg-CS) protein repeat epitopes of *Plasmodium falciparum* and of two rodent malaria agents, *Plasmodium berghei* and *Plasmodium yoelli*, were expressed in *Salmonella typhimurium* [130]. Immunization of mice with purified particles ensured not only specific B cell and T cell responses but also protection against a *P. berghei* challenge infection.

Second, a C-terminal segment (SR1) of SPAG-1, a sporozoite surface antigen of *Theileria annulata*, an infectious agent of cattle theileriosis, has been expressed as an MIR insertion [131]. The chimeric particles not only induced high titers of neutralizing antibodies, and a significant T cell response, but also showed some evidence of protection against sporozoite challenge [131], which allowed to recommend them for inclusion into future multicomponent vaccine [132].

Besides vaccine development. MIR insertions were used successfully for the development of anti-HEV immunoassays [133] and for mimicking of targeting moieties, or cell-receptor-recognizing sequences [134]. For the latter purpose, an RGD-containing epitope from the FMDV VP1 protein was exposed within the MIR, and the HBc-RGD chimeric particles not only elicited high levels of FMDV-neutralizing antibodies in guinea pigs, but also bound specifically to cultured eukaryotic cells, and to purified integrins [134].

Special interest is now devoted to construction of HBc display vectors with deletions of different length within the MIR. It is necessary to mention that some of the MIR insertions, which have been reviewed above, carried short deletions within the MIR: aa 76–80 [104, 115–117], 79–81 [121], and 79–80 [122, 123]. Structural [59, 62] and numerous experimental [119, 120, 135, 136] data convinced us that the region between the two conserved glycines G73 and G94 can be used as a target for deletions, rearrangements, and substitutions. For optimal immunogenicity of the insert, it is extremely important that deletions of proper aa residues within this region abrogate the intrinsic HBc antigenicity/immunogenicity [135, 136]. Besides the ability of the HBc carrier moiety to provide T cell help to inserted preS1 sequences, HBc carrier ensures the T-cell-independent character of humoral response against inserted epitopes in the MIR-deleted variants as well [96].

Taking into account the unique properties of the HBc carrier, the natural HBc deletion variants (e.g. 86–93 and

77–93) occurring in patients with progressive liver disease may deserve special attention as new carrier candidates [137]. However, insertion of 45 N-terminal aa of hantavirus nucleocapsid protein between aa 86 and 93 of HBc abolished the formation of chimeric VLPs [Koletzki. Preikschat. Meisel and Ulrich, unpubl. data].

An attempt to replace a more expanded fragment of the MIR, aa 72–89 within the HBcΔ by the HEV capsid epitope of 42 aa led to the production of capsomere-like 12-nm particles presumably constituted by the assembly of six dimers of the HBc protein [133].

In addition to empirical methods, dependence of self-assembly capabilities of MIR-inserted HBc proteins upon hydrophobicity, volume and other chemical properties of insertions was studied by computer calculations [138, 139].

Therefore, internal insertions into the HBc carrier offered strong possibilities of providing foreign epitope insertions with B cell and T cell immunologic activity. Although an attempt to insert a CTL epitope into the MIR of the HBc molecule was unsuccessful [129], the ability of the internal HBc vectors to support CTL activities must be explored further.

C-Terminal Insertions

Regarding the C-terminal insertions. HBc positions 144, 149, and 156 were used most frequently as target sites for foreign insertions. The capacity of the constructed vectors usually exceeded 100 aa residues, depending on the structure of insertion. The C-terminal insertions involved two types of vectors, encoding either full-length or C-terminally truncated HBc. In spite of the fact that capsids formed by the C-terminally truncated HBc derivatives (HBcΔ) are usually less stable than the capsids formed by full-length HBc proteins, high-level synthesis in bacteria and dissociation/reassociation capabilities of the HBcΔ are advantageous. Moreover, foreign insertions at the C-terminus can exert a stabilizing effect on chimeric HBcΔ derivatives, especially if internal insertions are introduced at the same construct [Borisova et al., in preparation]. In some cases, the inserted sequences are exposed, at least partially, at the surface of the HBc particle, but their specific B cell immunogenicity is usually low.

Full-length HBc vectors were used for insertion of fragments from the HBV preS [140–142], HIV-1 gp41 [143, 144], and simian immunodeficiency virus Env [100]. Further, expression by vaccinia virus of chimeric HBc carrying the long immediate-early CTL epitope from pp89 protein of murine cytomegalovirus (MCMV) at HBc position

179 led to induction of T-lymphocyte-mediated protec-
tive immunity against lethal MCMV infection [145].

C-terminally truncated vectors ensured high level of
synthesis and excellent self-assembly, but only moderate
specific immunogenicity of the inserted epitopes. These
vectors were used for expression of epitopes from the
HBV preS [44, 104, 140–142, 146], and HBsAg [12, 146,
147] regions: HIV-1 gp120 [120, 122, 123, 148], gp41
[146], Gag [148, 149], and Nef [143]; bovine leukemia
virus gp51 [140, 150]; human cytomegalovirus gp58
[108]; hantavirus nucleocapsid [113, 114, 151, 152], and
HEV capsid [133]. Although the fusion of 45 aa of the
Puumala hantavirus nucleocapsid protein allowed the for-
mation of chimeric VLPs, they were unable to induce a
protective immune response in the bank vole animal
model [109]. Chimeric HBc particles carrying C-terminal-
ly two virus-neutralizing epitopes from the FMDV VP1
(200–213, 131–160) showed excellent capability to self-
assemble, but failed to protect animals against FMDV
infection [153]. Finally, C-terminal insertions of the HCV
core protein demonstrated the extraordinary capacity of
the HBc particle as a VLP carrier: a 559 aa long insertion
did not prevent self-assembly of chimeras, and even 741
aa long insertion allowed production and self-assembly of
chimeras to some extent [154]. C-terminally added HCV
core [155, 156] and NS3 [157] sequences were used suc-
cessfully for detection of specific antibodies in HCV
enzyme immunoassay.

Although C-terminal additions have not met with suc-
cess in terms of induction of antibodies, a new attempt
was undertaken to insert a conserved sequence of 47 aa
residues from several proteins of Porphyromonas gingi-
valis [158]. Although in this case the chimeric particles
purified from E. coli were recognized by the host's im-
mune system and induced specific antibodies, they did
not protect mice against bacterial challenge.

Very recently, a 17-kD nuclease was packaged into the
interior of HBc capsids after fusion to the HBc position
155 [159]. The packaged nuclease retained enzymatic
activity, and the chimeric protein was able to form mosaic
particles with the wild-type HBc protein.

Special Applications of the HBc Particle as an Epitope Carrier

Scanning of the VLP Carrier-Encoding Gene

'Scanning' of the gene encoding the putative VLP sub-
unit by a short epitope as an immunological marker, in
order to find out gene regions, which are indifferent for

foreign insertions, but can ensure desirable structural
and/or immunological behavior of the latter [103, 119,
136, 151]. For this purpose, a short HBV preS1 epitope
31-DPAFRA-36 (or DPAFR, or DPAF) necessary and
sufficient [160] to be recognized by monoclonal antibody
MA18/7 [161] has been used. The behavior of the
DPAFR epitope was systematically compared after intro-
duction into all preferred insertion sites of the HBc mole-
cule at positions 2, 78, 144, and 183 [103], and into the
MIR carrying deletions of different length [136].

Mosaic Particles

A strategy to construct mosaic particles was based on
the introduction of a linker containing translational stop
codons (UGA, or UAG) between sequences encoding a
C-terminally truncated HBcΔ and a foreign protein se-
quence [103, 110, 114; 152, 162]. Expression of such
recombinant gene in an E. coli suppressor strain leads to
the simultaneous synthesis of both HBcΔ as a helper moi-
ety and a read-through fusion protein containing a foreign
sequence. This technology allowed incorporation into,
and presentation onto mosaic particles of 45 [109], 114
[114], 120 [110], and even 213 [Kazaks et al., in prepara-
tion] aa long segments of hantavirus nucleocapsid, al-
though nonmosaic HBcΔ carrying the hantavirus segment
at the C-terminus were unable to self-assemble. However,
in the animal model mosaic particles carrying 45 and 114
aa of the hantavirus nucleocapsid protein failed to induce
or induced only a marginal protective response [109,
114].

Easy Purification of Chimeric HBc Particles

Important practical advantage of the HBc model con-
sists in the fact that chimeric HBc-derived particles are
easy to purify by gel filtration or sucrose gradient centrifu-
gation, because of their particulate nature. C-terminally
truncated variants can be subjected to dissociation with
subsequent re-association, in order to remove internal
impurities and produce nucleic acid-free preparations. A
special purification protocol for preparation of HBc de-
rivatives of vaccine quality was elaborated by addition of
a 6 histidine tag to the truncated C-terminus of the HBc
protein [163]. On the other hand, the ability of full-length
or special chimeric HBc derivatives to controlled encapsi-
dation of nucleic acids may be used for the further devel-
opment of this carrier for gene therapy experiments.

Chim ric HBc Particles as a R al Vaccine and Gen Therapy Tool

Application of Chimeric HBc Particles as Vaccine Candidates

In spite of their status as an inner antigen, unmodified natural HBc particles were found in the middle 80s to be able to provide protection against HBV infection in chimpanzees [164–166]. The attempts to include HBc particles into HBV vaccines by retroviral [167] and DNA-based [168] expression or as a CTL epitope 18–27 of the HBc [169, 170] are now in progress.

In woodchucks, HBc particles were shown to protect animals after immunization, probably via T-cell mechanisms, since antibodies were not important for this protection [171]. Further interest to protective capabilities of the HBc was inspired by successful protection of woodchucks from WHV infection with the major WHc T-epitope peptide 97–110 [172].

The Celltech-Medeva company started recently a Hepacore project, which is oriented onto construction of therapeutic vaccines on the basis of chimeric HBc derivatives. A study on healthy volunteers using chimeric HBc particles containing the preS1 sequence 20–47 inserted into the MIR is planned for the third quarter of 2000. This study will evaluate the safety and tolerability of the Hepacore product, as well as provide immune response information that will be useful in designing further trials aimed at examining its potential in the immunotherapy of patients chronically infected with HBV [M. Page, pers. commun.].

Remarkable success in the movement to real HBc-based vaccines was achieved recently by construction of the HBc-M2 chimeras [111]. Due to the conserved nature of the M2 protein sequence, the HBc-M2 vaccine promises broad-spectrum, long-lasting protection against influenza A infections.

Strong hantavirus-neutralizing activity was shown also in the case of internal insertions of the hantavirus nucleocapsid epitopes [113, 114]. Further, chimeric HBc particles were generated carrying aa 1–45 and aa 75–119 of hantavirus nucleocapsid protein at aa position 78 and behind aa 144 of HBcΔ [113]. However, the combination of the major protective region of the nucleocapsid protein located between aa 1–45 and a second minor protective region did not improve the protective potential in the animal model, when compared to the particles carrying only the first region [109].

Immunization of mice with HBc-CS particles, which were expressed in and purified from S. typhimurium,

ensured not only specific B cell and T cell responses, bu also protection against a P. berghei challenge infectior [130, 173–176]. In general, expression of the recombinan genes in S. typhimurium suggested the promising idea o: generation of oral vaccines on the basis of live, avirulent strains of Salmonella species [9, 44, 117. 118, 128, 130. 173–176]. Thus, the efficacy of a single oral immunization of BALB/c mice with a recombinant S. typhimurium carrying an HBc-preS [174] and HBc-CS [175] chimera has been shown. In this case, the HBc-preS chimera contained aa 27–53 of the preS1 between positions 75 and 81 of the HBc protein and aa 133–143 of the preS2 fused C-terminally to position 156 of the HBc protein [104, 117]. However, volunteers that received oral Salmonella HBc-preS vaccine failed to develop humoral and cellular responses to hepatitis B antigens [177].

The chimeric HBc-SR1 particles, carrying a segment of a SPAG-1 of T. annulata is also regarded as a potential sporozoite vaccine challenge [131, 132].

Very recently, promising Salmonella expression variant of HBc-derived chimeras was achieved with internally inserted HBs 'a' epitope [178, 179]. A single rectal immunization with this HBc-HBs recombinant induced humoral and cellular immune response to HBc and HBs, and formation of specific mucosal immunity [179].

Furthermore, the following HBc derivatives were reported as infectious agent-neutralizing and potentially protective: HBc chimeras carrying N-terminal insertions of epitopes of FMDV VP1 [1, 2, 97] and outer membrane protein P.69 (pertactin) from B. pertussis [101], and internal insertion of the HRV-2 VP2 epitope [98, 99].

Chimeric HBc Particles for Gene Therapy?

The latest advances in the field show that chimeric VLPs are capable to present not only immunological epitopes but also other functional protein motifs, such as DNA and/or RNA binding and packaging sites, receptors and receptor binding sequences, immunoglobulins, elements recognizing low molecular mass substrates. It moves inevitably the VLP ideology from the conventional area of vaccine and diagnostic tool design to genetic vaccine and therapy applications.

The use of chimeric HBc particles for gene therapy requires two necessary capabilities, to pack the DNA or RNA genes of interest and to be taken up by target cells. It is shown experimentally that RNA may be packaged in vivo and in vitro by natural HBc particles [54], as well as by their derivatives carrying short DNA/RNA packaging sequences [Borisova et al., pers. commun.]. HBc particles with changed C-terminal part of the HBc molecule offer

prospects of nucleic acid packaging in vitro by a simple disassembly/reassembly procedure. Can natural HBc particles attach to eukaryotic cells and enable the uptake of the incorporated nucleic acid material? If yes, the further steps of intracellular expression of the incorporated nucleic acids can be accomplished by existing mechanisms [58].

For precise targeting, chimeric HBc particles must be provided with specific addresses, which may recognize appropriate receptors on special types of eukaryotic cells. The principal possibility of this approach was shown by construction of HBc-RGD particles [134].

Although the intracellular fate of the internalized HBc particles and their uptake mechanism remain still unclear, the chimeric HBc derivatives provided with the receptor-recognizing addresses and NA-packaging motifs may possibly become useful tools for gene delivery into a wide variety of cells.

References

1 Newton SE, Clarke BE, Appleyard G, Francis MJ, Carroll AR, Rowlands DJ, Skehel J, Brown F: New approaches to FMDV antigen presentation using vaccinia virus; in Chanock RM, Lerner RA, Brown F, Ginsberg H (eds): Vaccines 87. Modern approaches to new vaccines: Prevention of AIDS and other viral, bacterial and parasitic diseases. Cold Spring Harbor Laboratory. Cold Spring Harbor, 1987, pp 12–21.

2 Clarke BE, Newton SE, Carroll AR, Francis MJ, Appleyard G, Syred AD, Highfield PE, Rowlands DJ, Brown F; Improved immunogenicity of a peptide epitope after fusion to hepatitis B core protein. Nature 1987;330:381–384.

3 Borisova G, Bundule M, Grinstein E, Dreilina D, Dreimane A, Kalis J, Kozlovskaya T, Loseva V, Ose V, Pumpen P, Pushko P, Snikere D, Stankevica E, Tsibinogin V, Gren EJ: Recombinant capsid structures for exposure of protein antigenic epitopes. Mol Gen (Life Sci Adv) 1987;6:169–174.

4 Gerlich WH, Bruss V: Functions of hepatitis B virus proteins and molecular targets for protective immunity; in Ellis RW (ed): Hepatitis B Vaccines in Clinical Practice. New York, Dekker, 1993, pp 41–82.

5 Nassal M, Schaller H: Hepatitis B virus nucleocapsid assembly; in Doerfler W, Bohm P (eds): Virus Strategies. Molecular Biology and Pathogenesis. Weinheim, VCH, 1993, pp 41–75.

6 Kann M, Gerlich WH: Hepatitis B; in Collier L, Balows A, Sussman M (eds): Topley & Wilson's Microbiology and Microbial Infections. London, Arnold, 1998, pp 745–774.

7 Milich DR, Peterson DL, Zheng J, Hughes JL, Wirtz R, Schodel F: The hepatitis nucleocapsid as a vaccine carrier moiety. Ann NY Acad Sci 1995;754:187–201.

8 Pumpens P, Borisova GP, Crowther RA, Grens E: Hepatitis B virus core particles as epitope carriers. Intervirology 1995;38:63–74.

9 Schödel F, Peterson D, Hughes J, Wirtz R, Milich D: Hybrid hepatitis B virus core antigen as a vaccine carrier moiety. I. Presentation of foreign epitopes. J Biotechnol 1996;44:91–96.

10 Ulrich R, Nassal M, Meisel H, Krüger DH: Core particles of hepatitis B virus as carrier for foreign epitopes. Adv Virus Res 1998;50:141–182.

11 Pumpens P, Grens E: Hepatitis B core particles as a universal display model: A structure-function basis for development. FEBS Lett 1999;442:1–6.

12 Murray K, Shiau AL: The core antigen of hepatitis B virus as a carrier for immunogenic peptides. Biol Chem 1999;380:277–283.

13 Hirschman SZ, Price P, Garfinkel E, Christman J, Acs G: Expression of cloned hepatitis B virus DNA in human cell cultures. Proc Natl Acad Sci USA 1980;77:5507–5511.

14 Gough NM, Murray K: Expression of the hepatitis B virus surface, core and E antigen genes by stable rat and mouse cell lines. J Mol Biol 1982;162:43–67.

15 Gough NM: Core and E antigen synthesis in rodent cells transformed with hepatitis B virus DNA is associated with greater than genomic length viral messenger RNAs. J Mol Biol 1983;165:683–699.

16 Will H, Cattaneo R, Pfaff E, Kuhn C, Roggendorf M, Schaller H: Expression of hepatitis B antigens with a simian virus 40 vector. J Virol 1984;50:335–342.

17 Roossinck MJ, Jameel S, Loukin SH, Siddiqui A: Expression of hepatitis B viral core region in mammalian cells. Mol Cell Biol 1986;6:1393–1400.

18 Weimer T, Salfeld J, Will H: Expression of the hepatitis B virus core gene in vitro and in vivo. J Virol 1987;61:3109–3113.

19 McLachlan A, Milich DR, Raney AK, Riggs MG, Hughes JL, Sorge J, Chisari FV: Expression of hepatitis B virus surface and core antigens: Influences of pre-S and precore sequences. J Virol 1987;61:683–692.

20 Schlicht HJ, Schaller H: The secretory core protein of human hepatitis B virus is expressed on the cell surface. J Virol 1989;63:5399–5404.

21 Kunke D, Broucek J, Kutinova L, Nemeckova S, Ludvikova V, Stmad I, Kramosil J, Nemcova J, Schramlova J, Simonova V, et al: Vaccinia virus recombinants co-expressing hepatitis B virus surface and core antigens. Virology 1993;195:132–139.

22 Jean-Jean O, Levrero M, Will H, Perricaudet M, Rossignol JM: Expression mechanism of the hepatitis B virus (HBV) C gene and biosynthesis of HBe antigen. Virology 1989;170:99–106.

23 Standring DN, Ou JH, Masiarz FR, Rutter WJ: A signal peptide encoded within the precore region of hepatitis B virus directs the secretion of a heterogeneous population of e antigens in Xenopus oocytes. Proc Natl Acad Sci USA 1988;85:8405–8409.

24 Takehara K, Ireland D, Bishop DHL: Co-expression of the hepatitis B surface and core antigens using baculovirus multiple expression vectors. J Gen Virol 1988;69:2763–2777.

25 Lanford RE, Notvall L: Expression of hepatitis B virus core and precore antigens in insect cells and characterization of a core-associated kinase activity. Virology 1990;176:222–233.

26 Hilditch CM, Rogers LJ, Bishop DH: Physicochemical analysis of the hepatitis B virus core antigen produced by a baculovirus expression vector. J Gen Virol 1990;71:2755–2759.

27 Seifer M, Hamatake R, Bifano M, Standring DN: Generation of replication-competent hepatitis B virus nucleocapsids in insect cells. J Virol 1998;72:2765–2776.

28 Miyanohara A, Imamura T, Araki M, Sugawara K, Ohtomo N, Matsubara K: Expression of hepatitis B virus core antigen in Saccharomyces cerevisiae: Synthesis of two polypeptides translated from different initiation codons. J Virol 1986;59:176–180.

29 Kniskern PJ, Hagopian A, Montgomery DL, Burke P, Dunn NR, Hofmann KJ Miller WJ, Ellis RW: Unusually high-level expression of a foreign gene (hepatitis B virus core antigen) in Saccharomyces cerevisiae. Gene 1986;46:135–141.

30 Imamura T, Sugahara K, Adachi S, Miyatsu Y, Mizokami H, Matsuaka T: Purification and characterization of the hepatitis B virus core antigen produced in the yeast Saccharomyces cerevisiae. J Biotechnol 1988;8:149–162.

31 Shiosaki K, Takata K, Nishimura S, Mizokami H, Matsubara K: Production of hepatitis B virion-like particles in yeast. Gene 1991;106:143–149.

32 Tsuda S, Yoshioka K, Tanaka T, Iwata A, Yoshikawa A, Watanabe Y, Okada Y: Application of the human hepatitis B virus core antigen from transgenic tobacco plants for serological diagnosis. Vox Sang 1998;74:148–155.

33 Burrell CJ, Mackay P, Greenaway PJ, Hofschneider PH, Murray K: Expression in Escherichia coli of hepatitis B virus DNA sequences cloned in plasmid pBR322. Nature 1979;279:43–47.

34 Pasek M, Goto T, Gilbert W, Zink B, Schaller H, MacKay P, Leadbetter G, Murray K: Hepatitis B virus genes and their expression in E. coli. Nature 1979;282:575–579.

35 Edman JC, Hallewell RA, Valenzuela P, Goodman HM, Rutter WJ: Synthesis of hepatitis B surface and core antigens in E. coli. Nature 1981;291:503–506.

36 Stahl S, MacKay P, Magazin M, Bruce SA, Murray K: Hepatitis B virus core antigen: Synthesis in Escherichia coli and application in diagnosis. Proc Natl Acad Sci USA 1982;79:1606–1610.

37 Borisova GP, Pumpen PP, Bichko VV, Pushko PM, Kalis JV, Dishler AV, Gren EJ, Tsibinogin VV, Kukaine RA: Structure and expression in Escherichia coli cells of the core antigen gene of the human hepatitis B virus (in Russian). Dokl Akad Nauk SSSR 1984;279:1245–1249.

38 Uy A, Bruss V, Gerlich WH, Kochel HG, Thomssen R: Precore sequence of hepatitis B virus inducing e antigen and membrane association of the viral core protein. Virology 1986;155:89–96.

39 Lanford RE, Notvall LM, Dreesman GR, Harrison CR, Lockwood D, Burk KH: Expression and characterization of hepatitis B virus pre-core-core antigen in E. coli. Viral Immunol 1987;1:97–109.

40 Nassal M: Total chemical synthesis of a gene for hepatitis B virus core protein and its functional characterization. Gene 1988;66:279–294.

41 Khudyakov YuE, Kalinina TI, Neplyueva VS, Gazina EV, Kadoshnikov YuP, Bogdanova SL, Smirnov VD: The effect of the structure of the terminal regions of the hepatitis B virus gene C polypeptide on the formation of core antigen (HBcAg) particles. Biomed Sci 1991;2:257–265.

42 Maassen A, Rehfeldt A, Kiessig S, Ladhoff A, Hohne WE, Meisel H: Comparison of three different recombinant hepatitis B virus core particles expressed in Escherichia coli. Arch Virol 1994;135:131–142.

43 Hardy K, Stahl S, Kupper H: Production in B. subtilis of hepatitis B core antigen and of major antigen of foot and mouth disease virus. Nature 1981;293:481–483.

44 Schödel F, Milich DR, Will H: Hepatitis B virus nucleocapsid/pre-S2 fusion proteins expressed in attenuated Salmonella for oral vaccination. J Immunol 1990;145:4317–4321.

45 Schröder R, Maassen A, Lippoldt A, Borner T, von Baehr R, Dobrowolski P: Expression of the core antigen gene of hepatitis B virus (HBV) in Acetobacter methanolicus using broad host-range vectors. Appl Microbiol Biotechnol 1991;35:631–637.

46 Cohen BJ, Richmond JE: Electron microscopy of hepatitis B core antigen synthesized in E. coli. Nature 1982;296:677–679.

47 Yamaguchi M, Hirano T, Sugahara K, Mizokami H, Araki M, Matsubara K: Electron microscopy of hepatitis B virus core antigen expressing yeast cells by freeze-substitution fixation. Eur J Cell Biol 1988;47:138–143.

48 Kenney JM, von Bonsdorff CH, Nassal M, Fuller SD: Evolutionary conservation in the hepatitis B virus core structure: Comparison of human and duck cores. Structure 1995;3:1009–1019.

49 Scaglioni PP, Melegari M, Wands JR: Posttranscriptional regulation of hepatitis B virus replication by the precore protein. J Virol 1997;71:345–353.

50 Kann M, Gerlich WH: Replication of hepatitis B virus; in Harrison TJ, Zuckerman AJ (eds): Molecular Medicine of Viral Hepatitis. Chichester, Wiley, 1997, pp 63–87.

51 Dyson MR, Murray K: Selection of peptide inhibitors of interactions involved in complex protein assemblies: Association of the core and surface antigens of hepatitis B virus. Proc Natl Acad Sci USA 1995;92:2194–2198.

52 Poisson F, Severac A, Hourioux C, Goudeau A, Roingeard P: Both pre-S1 and S domains of hepatitis B virus envelope proteins interact with the core particle. Virology 1997;228:115–120.

53 Böttcher B, Tsuji N, Takahashi H, Dyson N, Zhao S, Crowther RA, Murray K: Peptides that block hepatitis B virus assembly: Analysis by cryomicroscopy, mutagenesis and transfection. EMBO J 1998;17:6839–6845.

54 Kann M, Gerlich WH: Effect of core protein phosphorylation by protein kinase C on encapsidation of RNA within core particles of hepatitis B virus. J Virol 1994;68:7993–8000.

55 Liao W, Ou JH: Phosphorylation and nuclear localization of the hepatitis B virus core protein: Significance of serine in the three repeated SPRRR motifs. J Virol 1995;69:1025–1029.

56 Lan YT, Li J, Liao W, Ou J: Roles of the three major phosphorylation sites of hepatitis B virus core protein in viral replication. Virology 1999;259:342–348.

57 Kann M, Bischof A, Gerlich WH: In vitro model for the nuclear transport of the hepadnavirus genome. J Virol 1997;71:1310–1316.

58 Kann M, Sodeik B, Vlachou A, Gerlich WH, Helenius A: Phosphorylation-dependent binding of hepatitis B virus core particles to the nuclear pore complex. J Cell Biol 1999;145:45–55.

59 Crowther RA, Kiselev NA, Böttcher B, Berriman JA, Borisova GP, Ose V, Pumpens P: Three dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. Cell 1994;77:943–950.

60 Böttcher B, Wynne SA, Crowther RA: Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. Nature 1997;386:88–91.

61 Conway JF, Cheng N, Zlotnick A, Wingfield PT, Stahl SJ, Steven AC: Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. Nature 1997;386:91–94.

62 Wynne SA, Crowther RA, Leslie AG: The crystal structure of the human hepatitis B virus capsid. Mol Cell 1999;3:771–780.

63 Borisova GP, Kalis JV, Pushko PM, Tsibinogin VV, Loseva VJ, Ose VP, Stankevica EI, Dreimane AJ, Snikere DJ, Grinstein EE, Pumpen PP, Gren EJ: Genetically engineered mutants of the core antigen of the human hepatitis B virus preserving the ability for native self-assembly (in Russian). Dokl Akad Nauk SSSR 1988;298:1474–1478.

64 Gallina A, Bonelli F, Zentilin L, Rindi G, Muttini M, Milanesi G: A recombinant hepatitis B core antigen polypeptide with the protamine-like domain deleted self-assembles into capsid particles but fails to bind nucleic acids. J Virol 1989;63:4645–4652.

65 Inada T, Misumi Y, Seno M, Kanezaki S, Shibata Y, Oka Y, Onda H: Synthesis of hepatitis B virus e antigen in E. coli. Virus Res 1989;14:27–47.

66 Melegari M, Bruss V, Gerlich WH: The arginine-rich carboxy-terminal domain is necessary for RNA packaging by hepatitis B virus core protein; in Hollinger FB, Lemon SM, Margolis HS (eds): Viral Hepatitis and Liver Disease. Baltimore, Williams & Wilkins, 1991, pp 164–168.

67 Birnbaum F, Nassal M: Hepatitis B virus nucleocapsid assembly: Primary structure requirements in the core protein. J Virol 1990; 64:3319-3330.

68 Bundule MA, Bichko VV, Saulitis JB, Liepins EE, Borisova GP, Petrovskis IA, Tsibinogin VV, Pumpen PP, Gren EJ: C-terminal polyarginine tract of hepatitis B core antigen is located on the outer capsid surface (in Russian). Dokl Akad Nauk SSSR 1990:312:993-996.

69 Seifer M, Standring DN: Assembly and antigenicity of hepatitis B virus core particles. Intervirology 1995:38:47-62.

70 Zlotnick A, Cheng N, Conway JF, Booy FP, Steven AC, Stahl SJ, Wingfield PT: Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. Biochemistry 1996:35:7412-7421.

71 Al-Khayat HA, Bhella D, Kenney JM, Roth JF, Kingsman AJ, Martin-Rendon E, Saibil HR: Yeast Ty retrotransposons assemble into viruslike particles whose T-numbers depend on the C-terminal length of the capsid protein. J Mol Biol 1999:292:65-73.

72 Hoofnagle JH, Gerety RJ, Barker LF: Antibody to hepatitis-B-virus core in man. Lancet 1973: 2:869-873.

73 Chisari FV, Ferrari C: Hepatitis B virus immunopathogenesis. Annu Rev Immunol 1995;13: 29-60.

74 Milich DR, McLachlan A: The nucleocapsid of hepatitis B virus is both a T-cell-independent and a T-cell-dependent antigen. Science 1986: 234:1398-1401.

75 Milich DR, Peterson DL, Schodel F, Jones JE, Hughes JL: Preferential recognition of hepatitis B nucleocapsid antigens by Th1 or Th2 cells is epitope and major histocompatibility complex dependent. J Virol 1995;69:2776-2785.

76 Milich DR, Schodel F, Hughes JL, Jones JE, Peterson DL: The hepatitis B virus core and e antigens elicit different Th cell subsets: Antigen structure can affect Th cell phenotype. J Virol 1997;71:2192-2201.

77 Milich DR, McLachlan A, Thornton GB, Hughes JL: Antibody production to the nucleocapsid and envelope of the hepatitis B virus primed by a single synthetic T cell site. Nature 1987;329:547-549.

78 Milich DR, Chen M, Schodel F, Peterson DL, Jones JE, Hughes JL: Role of B cells in antigen presentation of the hepatitis B core. Proc Natl Acad Sci USA 1997;94:14648-14653.

79 Mondelli M, Vergani GM, Alberti A, Vergani D, Portmann B, Eddleston AL, Williams R: Specificity of T lymphocyte cytotoxicity to autologous hepatocytes in chronic hepatitis B virus infection: Evidence that T cells are directed against HBV core antigen expressed on hepatocytes. J Immunol 1982;129:2773-2778.

80 Rehermann B, Ferrari C, Pasquinelli C, Chisari FV: The hepatitis B virus persists for decades after patients' recovery from acute viral hepatitis despite active maintenance of a cytotoxic T-lymphocyte response. Nat Med 1996:2:1104-1108.

81 Salfeld J, Pfaff E, Noah M, Schaller H: Antigenic determinants and functional domains in core antigen and e antigen from hepatitis B virus. J Virol 1989:63:798-808.

82 Sallberg M, Ruden U, Magnius LO, Harthus HP, Noah M, Wahren B: Characterisation of a linear binding site for a monoclonal antibody to hepatitis B core antigen. J Med Virol 1991: 33:248-252.

83 Sallberg M, Ruden U, Wahren B, Noah M, Magnius LO: Human and murine B-cells recognize the HBeAg/β (or HBe2) epitope as a linear determinant. Mol Immunol 1991:28:719-726.

84 Sallberg M, Pushko P, Berzinsh I, Bichko V, Sillekens P, Noah M, Pumpens P, Grens E, Wahren B, Magnius LO: Immunochemical structure of the carboxy-terminal part of hepatitis B e antigen: Identification of internal and surface-exposed sequences. J Gen Virol 1993: 74:1335-1340.

85 Ferrari C, Bertoletti A, Penna A, Cavalli A, Valli A, Missale G, Pilli M, Fowler P, Giuberti T, Chisari FV, Fiaccadori F: Identification of immunodominant T cell epitopes of the hepatitis B virus nucleocapsid antigen. J Clin Invest 1991:88:214-222.

86 Wakita T, Kakumu S, Tsutsumi Y, Yoshioka K, Machida A, Mayumi M: Gamma-interferon production in response to hepatitis B core protein and its synthetic peptides in patients with chronic hepatitis B virus infection. Digestion 1990:47:149-155.

87 Diepolder HM, Jung MC, Wierenga E, Hoffmann RM, Zachoval R, Gerlach TJ, Scholz S, Heavner G, Riethmuller G, Pape GR: Anergic TH1 clones specific for hepatitis B virus (HBV) core peptides are inhibitory to other HBV corespecific CD4+ T cells in vitro. J Virol 1996:70: 7540-7548.

88 Milich DR, McLachlan A, Moriarty A, Thornton GB: Immune response to hepatitis B virus core antigen (HBcAg): Localization of T cell recognition sites within HBcAg/HBeAg. J Immunol 1987:139:1223-1231.

89 Milich DR, Hughes JL, Houghten R, McLachlan A, Jones JE: Functional identification of agretopic and epitopic residues within an HBcAg T cell determinant. J Immunol 1989; 143:3141-3147.

90 Bertoletti A, Chisari FV, Penna A, Guilhot S, Galati L, Missale G, Fowler P, Schlicht HJ, Vitiello A, Chesnut RC, Fiaccadori F, Ferrari C: Definition of a minimal optimal cytotoxic T-cell epitope within the hepatitis B virus nucleocapsid protein. J Virol 1993:67:2376-2380.

91 Missale G, Redeker A, Person J, Fowler P, Guilhot S, Schlicht HJ, Ferrari C, Chisari FV: HLA-A31- and HLA-Aw68-restricted cytotoxic T cell responses to a single hepatitis B virus nucleocapsid epitope during acute viral hepatitis. J Exp Med 1993:177:751-762.

92 Tsai SL, Chen MH, Yeh CT, Chu CM, Lin AN, Chiou FH, Chang TH, Liaw YF: Purification and characterization of a naturally processed hepatitis B virus peptide recognized by CD8+ cytotoxic T lymphocytes. J Clin Invest 1996: 97:577-584.

93 Kuhober A, Pudollek HP, Reifenberg K, Chisari FV, Schlicht HJ, Reimann J, Schirmbeck R: DNA immunization induces antibody and cytotoxic T cell responses to hepatitis B core antigen in H-2b mice. J Immunol 1996:156: 3687-3695.

94 Kuhober A, Wild J, Pudollek HP, Chisari FV, Reimann J: DNA vaccination with plasmids encoding the intracellular (HBcAg) or secreted (HBeAg) form of the core protein of hepatitis B virus primes T cell responses to two overlapping Kb- and Kd-restricted epitopes. Int Immunol 1997:9:1203-1212.

95 Townsend K, Sallberg M, O'Dea J, Banks T, Driver D, Sauter S, Chang SM, Jolly DJ, Mento SJ, Milich DR, Lee WT: Characterization of CD8+ cytotoxic T-lymphocyte responses after genetic immunization with retrovirus vectors expressing different forms of the hepatitis B virus core and e antigens. J Virol 1997:71:3365-3374.

96 Fehr T, Skrastina D, Pumpens P, Zinkernagel RM: T cell-independent type I antibody response against B cell epitopes expressed repetitively on recombinant virus particles. Proc Natl Acad Sci USA 1998:95:9477-9481.

97 Beesley KM, Francis MJ, Clarke BE, Beesley JE, Dopping-Hepenstal PJ, Clare JJ, Brown F, Romanos MA: Expression in yeast of amino-terminal peptide fusions to hepatitis B core antigen and their immunological properties. Biotechnology (NY) 1990:8:644-649.

98 Clarke BE, Brown AL, Grace KG, Hastings GZ, Brown F, Rowlands DJ, Francis MJ: Presentation and immunogenicity of viral epitopes on the surface of hybrid hepatitis B virus core particles produced in bacteria. J Gen Virol 1990:71:1109-1117.

99 Brown AL, Francis MJ, Hastings GZ, Parry NR, Barnett PV, Rowlands DJ, Clarke BE: Foreign epitopes in immunodominant regions of hepatitis B core particles are highly immunogenic and conformationally restricted. Vaccine 1991:9:595-601.

100 Yon J, Rud E, Corcoran T, Kent K, Rowlands D, Clarke B: Stimulation of specific immune responses to simian immunodeficiency virus using chimeric hepatitis B core antigen particles. J Gen Virol 1992:73:2569-2575.

101 Charles IG, Li JL, Roberts M, Beesley K, Romanos M, Pickard DJ, Francis M, Campbell D, Dougan G, Brennan MJ, et al: Identification and characterization of a protective immunodominant B cell epitope of pertactin (P.69) from Bordetella pertussis. Eur J Immunol 1991:21:1147-1153.

102 Kalinina TI, Makeeva IV, Khudiakov IuF, Samoshin VV, Smirnova EA, Semiletov IuA, Pavliuchenkova RP, Kadoshnikov IuP, Smirnov VD: Introduction of heterologous epitopes at the N-terminal part of the hepatitis B core protein (in Russian). Mol Biol 1995:29: 199-210.

103 Lachmann S, Meisel H, Muselmann C, Koletzki D, Gelderblom HR, Borisova G, Kruger DH, Pumpens P, Ulrich R: Characterization of potential insertion sites in the core antigen of hepatitis B virus by the use of a short-sized model epitope. Intervirology 1999;42:51–56.

104 Schödel F, Moriarty AM, Peterson DL, Zheng JA, Hughes JL, Will H, Letureq DJ, McGee JS, Milich DR: The position of heterologous epitopes inserted in hepatitis B virus core particles determines their immunogenicity. J Virol 1992;66:106–114.

105 Moriarty AM, McGee JS, Winslow B, Inman D, Letureq DJ, Thornton GB, Hughes JL, Milich DR: Expression of HIV gag and env B-cell epitopes on the surface of HBV core particles and analysis of the immune responses generated to those epitopes; in Brown F, Chanock RM, Ginsberg HS, Lerner RA (eds): Vaccines 90. Cold Spring Harbor. Cold Spring Harbor Laboratory. 1990, pp 225–229.

106 Isaguliants MG, Kadoshnikov IuP, Kalinina TI, Khudiakov IuE, Semiletov IuA, Smirnov VD, Wahren B: Expression of HIV-1 epitopes included in particles formed by human hepatitis B virus nucleocapsid protein (in Russian). Biokhimiia 1996;61:532–545.

107 Isaguliants MG, Nordlund S, Sallberg M, Smirnov VD, Ruden U, Wahren B: HIV-1 epitopes exposed by hybrid hepatitis B core particles affect proliferation of peripheral blood mononuclear cells from HIV-1 positive donors. Immunol Lett 1996;52:37–44.

108 Tarar MR, Emery VC, Harrison TJ: Expression of a human cytomegalovirus gp58 antigenic domain fused to the hepatitis B virus nucleocapsid protein. FEMS Immunol Med Microbiol 1996;16:183–192.

109 Koletzki D, Lundkvist A, Brus Sjölander K, Gelderblom HR, Niedrig M, Meisel H, Kruger DH, Ulrich R: Puumala (PUU) hantavirus strain differences and insertion positions in the hepatitis B virus core antigen influence B-cell immunogenicity and protective potential of core-derived particles. Virology 2000; 276:364–375.

110 Koletzki D, Biel SS, Meisel H, Nugel E, Gelderblom HR, Krüger DH, Ulrich R: HBV core particles allow the insertion and surface exposure of the entire potentially protective region of Puumala hantavirus nucleocapsid protein. Biol Chem 1999;380:325–333.

111 Neirynck S, Deroo T, Saelens X, Vanlandschoot P, Jou WM, Fiers W: A universal influenza A vaccine based on the extracellular domain of the M2 protein. Nat Med 1999;5: 1157–1163.

112 Böttcher B, Dyson MR, Crowther RA: Finding the small difference: A nine amino acid extension to the hepatitis B core protein. Electron microscopy. 1998. ICEM14. Cancun. Mexico. 31 Aug to 4 Sept 1998. Symposium QQ. vol 1. pp 737–738.

113 Ulrich R, Lundkvist A, Meisel H, Koletzki D, Sjolander KB, Gelderblom HR, Borisova G, Schnitzler P, Darai G, Krüger DH: Chimaeric HBV core particles carrying a defined segment of Puumala hantavirus nucleocapsid protein evoke protective immunity in an animal model. Vaccine 1998;16:272–280.

114 Ulrich R, Koletzki D, Lachmann S, Lundkvist A, Zankl A, Kazaks A, Kurth A, Gelderblom HR, Borisova G, Meisel H, Krüger DH: New chimaeric hepatitis B virus core particles carrying hantavirus (serotype Puumala) epitopes: Immunogenicity and protection against virus challenge. J Biotechnol 1999;73: 141–153.

115 Kratz PA, Böttcher B, Nassal M: Native display of complete foreign protein domains on the surface of hepatitis B virus capsids. Proc Natl Acad Sci USA 1999;96:1915–1920.

116 Schödel F, Weimer T, Will H, Milich D: Recombinant HBV core particles carrying immunodominant B-cell epitopes of the HBV preS2-region; in Brown F, Chanock RM, Ginsberg HS, Lerner RA (eds): Vaccines 90. Cold Spring Harbor. Cold Spring Harbor Laboratory Press. 1990. pp 193–198.

117 Schödel F, Will H, Milich DR: Hybrid hepatitis-B virus core/pre-S particles expressed in live attenuated Salmonellae for oral immunization; in Brown F, Chanock RM, Ginsberg HS, Lerner RA (eds): Vaccines 91. Cold Spring Harbor. Cold Spring Harbor Laboratory Press. 1991. pp 319–325.

118 Schödel F, Peterson D, Hughes J, Milich DR: A virulent Salmonella expressing hybrid hepatitis B virus core/pre-S genes for oral vaccination. Vaccine 1993;11:143–148.

119 Borisova G, Borschukova Wanst O, Mezule G, Skrastina D, Petrovskis I, Dislers A, Pumpens P, Grens E: Spatial structure and insertion capacity of immunodominant region of hepatitis B core antigen. Intervirology 1996;39:16–22.

120 Borisova G, Borschukova O, Skrastina D, Mezule G, Dislers A, Petrovskis I, Ose V, Gusars I, Pumpens P, Grens E: Display vectors. 1. Hepatitis B core particle as a display moiety. Proc Latv Acad Sci 1997;51:1–7.

121 Clarke BE, Carroll AR, Brown AL, Jon J, Parry NR, Rud EW, Francis MJ, Rowlands DJ: Expression and immunological analysis of hepatitis-B core fusion particles carrying internal heterologous sequences: in Chanock RM, Ginsberg HS, Brown F, Lerner RA (eds): Vaccines 91. Cold Spring Harbor. Cold Spring Harbor Laboratory Press. 1991. pp 313–318.

122 von Brunn A, Reichhuber C, Brand M, Bechowsky B, Gurtler L, Eberle J, Kleinschmidt A, Erfle V, Schödel F: The principal neutralizing determinant (V3) of HIV-1 induces HIV-1-neutralizing antibodies upon expression on HBcAg particles; in Ginsberg HS, Brown F, Chanock RM, Lerner RA (eds): Vaccines 93. Cold Spring Harbor. Cold Spring Harbor Laboratory Press. 1993. pp 159–165.

123 von Brunn A, Brand M, Reichhuber C, Morys-Wortmann C, Deinhardt F, Schödel F: Principal neutralizing domain of HIV-1 is highly immunogenic when expressed on the surface of hepatitis B core particles. Vaccine 1993;11:817–824.

124 Borisova G, Arya B, Dislers A, Borschukova O, Tsibinogin V, Skrastina D, Eldarov MA, Pumpens P, Skryabin KG, Grens E: Hybrid hepatitis B virus nucleocapsid bearing an immunodominant region from hepatitis B virus surface antigen. J Virol 1993;67:3696–3701.

125 Makeeva IV, Kalinina TI, Khudiakov IuE, Samoshin VV, Smirnova EA, Semiletov IuA, Pavliuchenkova RP, Kadoshnikov IuP, Smirnov VD: Heterologous epitopes in the central part of the hepatitis B virus core protein (in Russian). Mol Biol 1995;29:211–224.

126 Loktev VB, Ilyichev AA, Eroshkin AM, Karpenko LI, Pokrovsky AG, Perebory AV, Svyatchenko VA, Ignat'ev GM, Smolina MI, Melamed NV, Lebedeva CD, Sandakhchiev LS: Design of immunogens as components of a new generation of molecular vaccines. J Biotechnol 1996;44:129–137.

127 Tindle RW, Herd K, Londono P, Fernando GJ, Chatfield SN, Malcolm K, Dougan G: Chimeric hepatitis B core antigen particles containing B- and Th-epitopes of human papillomavirus type 16 E7 protein induce specific antibody and T-helper responses in immunised mice. Virology 1994;200:547–557.

128 Londono LP, Chatfield S, Tindle RW, Herd K, Gao XM, Frazer I, Dougan G: Immunisation of mice using Salmonella typhimurium expressing human papillomavirus type 16 E7 epitopes inserted into hepatitis B virus core antigen. Vaccine 1996;14:545–552.

129 Street M, Herd K, Londono P, Doan T, Dougan G, Kast WM, Tindle RW: Differences in the effectiveness of delivery of B- and CTL-epitopes incorporated into the hepatitis B core antigen (HBcAg) c/e1-region. Arch Virol 1999;144:1323–1343.

130 Schödel F, Wirtz R, Peterson D, Hughes J, Warren R, Sadoff J, Milich D: Immunity to malaria elicited by hybrid hepatitis B virus core particles carrying circumsporozoite protein epitopes. J Exp Med 1994;180:1037–1046.

131 Boulter NR, Glass EJ, Knight PA, Bell-Sakyi L, Brown CG, Hall R: Theileria annulata sporozoite antigen fused to hepatitis B core antigen used in a vaccination trial. Vaccine 1995;13:1152–1160.

132 Boulter N, Brown D, Wilkie G, Williamson S, Kirvar E, Knight P, Glass E, Campbell J, Morzaria S, Nene V, Musoke A, d'Oliveira C, Gubbels MJ, Jongejan F, Hall R: Evaluation of recombinant sporozoite antigen SPAG-1 as a vaccine candidate against Theileria annulata by the use of different delivery systems. Trop Med Int Health 1999;4:A71–77.

133 Touze A, Enogat N, Buisson Y, Coursaget P: Baculovirus expression of chimeric hepatitis B virus core particles with hepatitis E virus epitopes and their use in a hepatitis E immunoassay. J Clin Microbiol 1999;37:438–441

134 Chambers MA. Dougan G. Newman J. Brown F, Crowther J. Mould AP. Humphries MJ. Francis MJ. Clarke B. Brown AL. Rowlands D: Chimeric hepatitis B virus core particles as probes for studying peptide-integrin interactions. J Virol 1996:70:4045–4052.

135 Borschukova O. Skrastina D. Dislers A. Petrovskis I. Ose V. Zamurujeva I. Borisova G: Modified hepatitis B core particles as possible vaccine carriers: in Brown F. Burton D. Doherty P. Mekalanos J. Norrby E (eds): Vaccines 97. Cold Spring Harbor. Cold Springer Harbor Laboratory Press. 1997. pp 33–37.

136 Borisova G. Borschukova O. Skrastina D. Dislers A. Ose V. Pumpens P. Grens E: Behavior of a short preS1 epitope on the surface of hepatitis B core particles. Biol Chem 1999: 380:315–324.

137 Preikschat P. Borisova G. Borschukova O. Dislers A. Mezule G. Grens E, Krüger DH. Pumpens P. Meisel H: Expression. assembly competence and antigenic properties of hepatitis B virus core gene deletion variants from infected liver cells. J Gen Virol 1999:80: 1777–1788.

138 Karpenko LI. Ivanisenko VA. Pika IS. Chikaev NA. Eroshkin AM. Melamed NV. Veremeiko TA. Ilyichev AA: Analysis of foreign epitope inserts in HBcAg. Approaches to solving the problem of core particle self-assembly (in Russian). Mol Biol (Moscow) 2000:34: 223–229.

139 Karpenko LI. Ivanisenko VA. Pika IA. Chikaev NA. Eroshkin AM. Veremeiko TA. Ilyichev AA: Insertion of foreign epitopes in HBcAg: How to make the chimeric particle assemble. Amino Acids 2000:18:329–337.

140 Borisova GP. Berzins I. Pushko PM. Pumpen P. Gren EJ. Tsibinogin VV. Loseva V. Ose V. Ulrich R. Siakkou H. Rosenthal HA: Recombinant core particles of hepatitis B virus exposing foreign antigenic determinants on their surface. FEBS Lett 1989:259:121–124.

141 Borisova GP. Berzins IG. Tsibinogin VV. Loseva VJ. Pushko PM. Ose VP. Dreilina DE. Pumpen PP. Gren EJ: Hepatitis B virus core antigen as a carrier for functionally active epitopes: Exposure of preS regions on the capsids (in Russian). Dokl Akad Nauk SSSR 1990: 312:751–754.

142 Zakis V. Skrastina D. Borisova G. Kuranova I: Immunodominance of T-cell epitopes on foreign sequences of the hepatitis B virus nucleocapsid fusion proteins; in Brown F. Chanock RM. Ginsberg HS. Lerner RA (eds): Vaccines 92. Cold Spring Harbor. Cold Spring Harbor Laboratory Press. 1992. pp 341–347.

143 Ulrich R. Meisel H. Soza A. Krüger DH. Ladholt A-M. Borisova G. Pumpen P: Characterization of chimeric core particles of HBV containing foreign epitopes; in Ginsberg HS. Brown F. Chanock RM. Lerner RA (eds): Vaccines 93 (Modern Approaches to New Vaccines including Prevention of AIDS). Cold Spring Harbor. Cold Springer Harbor Laboratory Press. 1993. pp 323–328.

144 Ulrich R. Borisova GP. Möhring R. Lätzsch I. Ose VP. Berzins IG. Dreilina DE. Pushko PM. Tsibinogin VV. Pumpen PP. Rosenthal HA. Gren EJ: Exposure of HIV-1 gp41 transmembrane protein epitopes on the surface of hepatitis B core antigen capsids (in Russian). Bioorg Khim 1990:16:1283–1285.

145 Del Val M. Schlicht HJ. Volkmer H. Messerle M. Reddehase MJ. Koszinowski UH: Protection against lethal cytomegalovirus infection by a recombinant vaccine containing a single nonameric T-cell epitope. J Virol 1991:65: 3641–3646.

146 Stahl SJ. Murray K: Immunogenicity of peptide fusions to hepatitis B virus core antigen. Proc Natl Acad Sci USA 1989:86:6283–6287.

147 Shiau AL. Murray K: Mutated epitopes of hepatitis B surface antigen fused to the core antigen of the virus induce antibodies that react with the native surface antigen. J Med Virol 1997:51:159–166.

148 Grene E. Mezule G. Borisova G. Pumpens P. Bentwich Z. Arnon R: Relationship between antigenicity and immunogenicity of chimeric hepatitis B virus core particles carrying HIV type 1 epitopes. AIDS Res Hum Retroviruses 1997:13:41–51.

149 Ulrich R. Borisova GP. Gren E. Berzin I. Pumpen P. Eckert R. Ose V. Siakkou H. Gren EJ. von Baehr R. Krüger DH: Immunogenicity of recombinant core particles of hepatitis B virus containing epitopes of human immunodeficiency virus 1 core antigen. Arch Virol 1992:126:321–328.

150 Ulrich R. Borisova GP. Siakkou H. Platzer C. Ose VP. Berzins IG. Dreilina DE. Pushko PM. Tsibinogin VV. Pumpen PP. Rosenthal HA. Gren EJ: Exposure of major immunodominant epitope of bovine leukemia virus envelope protein gp51 on the surface of hepatitis B core antigen capsids (in Russian). Mol Biol 1991:25:368–374.

151 Berzins I. Ulrich R. Meisel H. Krüger DH. Borisova G. Pumpens P. Grens E. Gelderblom H. Ladhoff A-M. Schnitzler P. Darai G. Bautz EKF: Sites in the core antigen of HBV allowing insertion of foreign epitopes: in Norrby E. Brown F. Chanock RM. Ginsberg HS (eds): Vaccines 94. Cold Spring Harbor. Cold Spring Harbor Laboratory Press. 1994. pp 301–308.

152 Ulrich R. Koletzki D. Zankl A. Schulz A. Meisel H. Krüger DH. Gelderblom HR. Dislers A. Borisova G. Pumpens P: A new strategy to generate mosaic HBcAg particles presenting foreign epitopes: in Brown F. Burton D. Doherty P. Mekalanos J. Norrby E (eds): Vaccines 97. Cold Spring Harbor. Cold Spring Laboratory Press. 1997. pp 235–240.

153 Nekrasova OV. Boichenko VE. Boldyreva EF. Borisova GP. Pumpen P. Perevozchikova NA. Korobko VG: Bacterial synthesis of immunogenic epitopes of foot-and-mouth disease virus fused either to human necrosis factor or to hepatitis B core antigen (in Russian). Bioorg Khim 1997:23:118–126.

154 Yoshikawa A. Tanaka T. Hoshi Y. Kato N. Tachibana K. Iizuka H. Machida A. Okamoto H. Yamasaki M. Miyakawa Y. Miyakawa Y. Mayumi M: Chimeric hepatitis B virus core particles with parts or copies of the hepatitis C virus core protein. J Virol 1993:67: 6064–6070.

155 Claeys H. Volckaerts A. De Beenhouwer H. Vermylen C: Association of hepatitis C virus carrier state with the occurrence of hepatitis C virus core antibodies. J Med Virol 1992:36: 259–264.

156 Wu CL. Leu TS. Chang TT. Shiau AL: Hepatitis C virus core protein fused to hepatitis B virus core antigen for serological diagnosis of both hepatitis C and hepatitis B infections by ELISA. J Med Virol 1999:57:104–110.

157 Claeys H. Volckaerts A. Mertens W. Liang Z. Fiten P. Opdenakker G: Localization and reactivity of an immunodominant domain in the NS3 region of hepatitis C virus. J Med Virol 1995:45:273–281.

158 Dawson JA. Macrina FL: Construction and immunologic evaluation of a Porphyromonas gingivalis subsequence peptide fused to hepatitis B virus core antigen. FEMS Microbiol Lett 1999:175:119–125.

159 Beterams G. Böttcher B. Nassal M: Packaging of up to 240 subunits of a 17kDa nuclease into the interior of recombinant hepatitis B virus capsids. FEBS Lett 2000:481:169–176.

160 Sominskaya I. Pushko P. Dreilina D. Kozlovskaya T. Pumpen P: Determination of the minimal length of preS1 epitope recognized by a monoclonal antibody which inhibits attachment of hepatitis B virus to hepatocytes. Med Microbiol Immunol 1992:181:215–226.

161 Heermann KH. Goldmann U. Schwartz U. Seyffarth T. Baumgarten H. Gerlich WH: Large surface proteins of hepatitis B virus containing the pre-S sequence. J Virol 1984: 52:396–402.

162 Koletzki D. Zankl A. Gelderblom HR. Meisel H. Dislers A. Borisova G. Pumpens P. Krüger DH. Ulrich R: Mosaic hepatitis B virus core particles allow insertion of extended foreign protein segments. J Gen Virol 1997:78:2049–2053.

163 Wizemann H. von Brunn A: Purification of E. coli-expressed HIS-tagged hepatitis B core antigen by Ni²⁺-chelate affinity chromatography. J Virol Methods 1999:77:189–197.

164 Iwarson S. Tabor E. Thomas HC. Snoy P. Gerety RJ: Protection against hepatitis B virus infection by immunization with hepatitis B core antigen. Gastroenterology 1985:88: 763–767.

165 Murray K. Bruce SA. Hinnen A. Wingfield P. van Erd PM. de Reus A. Schellekens H: Hepatitis B virus antigens made in microbial cells immunise against viral infection. EMBO J 1984:3:645–650.

166 Murray K. Bruce SA. Wingfield P. van Eerd P. de Reus A. Schellekens H: Protective immunisation against hepatitis B with an internal antigen of the virus. J Med Virol 1987:23: 101–107.

167 Sallberg M, Hughes J, Javadian A, Ronlov G, Hultgren C, Townsend K, Anderson CG, O'Dea J, Alfonso J, Eason R, Murthy KK, Jolly DJ, Chang SM, Mento SJ, Milich D, Lee WT: Genetic immunization of chimpanzees chronically infected with the hepatitis B virus, using a recombinant retroviral vector encoding the hepatitis B virus core antigen. Hum Gene Ther 1998;9:1719-1729.

168 Wild J, Gruner B, Metzger K, Kuhrober A, Pudollek HP, Hauser H, Schirmbeck R, Reimann J: Polyvalent vaccination against hepatitis B surface and core antigen using a dicistronic expression plasmid. Vaccine 1998;16:353-360.

169 Livingston BD, Crimi C, Grey H, Ishioka G, Chisari FV, Fikes J, Grey H, Chesnut RW, Sette A: The hepatitis B virus-specific CTL responses induced in humans by lipopeptide vaccination are comparable to those elicited by acute viral infection. J Immunol 1997;159:1383-1392.

170 Heathcote J, McHutchison J, Lee S, Tong M, Benner K, Minuk G, Wright T, Fikes J, Livingston B, Sette A, Chestnut R: A pilot study of the CY-1899 T-cell vaccine in subjects chronically infected with hepatitis B virus. The CY1899 T Cell Vaccine Study Group. Hepatology 1999;30:531-536.

171 Schödel F, Neckermann G, Peterson D, Fuchs K, Fuller S, Will H, Roggendorf M: Immunization with recombinant woodchuck hepatitis virus nucleocapsid antigen or hepatitis B virus nucleocapsid antigen protects woodchucks from woodchuck hepatitis virus infection. Vaccine 1993;11:624-628.

172 Menne S, Maschke J, Tolle TK, Lu M, Roggendorf M: Characterization of T-cell response to woodchuck hepatitis virus core protein and protection of woodchucks from infection by immunization with peptides containing a T-cell epitope. J Virol 1997;71:65-74.

173 Schödel F, Kelly SM, Peterson D, Milich D, Hughes J, Tinge S, Wirtz R, Curtiss R 3rd: Development of recombinant Salmonellae expressing hybrid hepatitis B virus core particles as candidate oral vaccines. Dev Biol Stand 1994;82:151-158.

174 Schödel F, Kelly SM, Peterson DL, Milich DR, Curtiss R 3rd: Hybrid hepatitis B virus core-pre-S proteins synthesized in avirulent Salmonella typhimurium and Salmonella typhi for oral vaccination. Infect Immun 1994;62:1669-1676.

175 Schödel F, Kelly S, Tinge S, Hopkins S, Peterson D, Milich D, Curtiss R 3rd: Hybrid hepatitis B virus core antigen as a vaccine carrier moiety. II. Expression in avirulent Salmonella spp. for mucosal immunization. Adv Exp Med Biol 1996;397:15-21.

176 Schödel F, Peterson D, Milich DR, Charoenvit Y, Sadoff J, Wirtz R: Immunization with hybrid hepatitis B virus core particles carrying circumsporozoite antigen epitopes protects mice against Plasmodium yoelii challenge. Behring Inst Mitt 1997;98:114-119.

177 Tacket CO, Kelly SM, Schödel F, Losonsky G, Nataro JP, Edelman R, Levine MM, Curtiss R 3rd: Safety and immunogenicity in humans of an attenuated Salmonella typhi vaccine vector strain expressing plasmid-encoded hepatitis B antigens stabilized by the Asd-balanced lethal vector system. Infect Immun 1997;65:3381-3385.

178 Karpenko LI, Il'ichev AA: Chimeric hepatitis B core antigen particles as a presentation system of foreign protein epitopes (in Russian). Vestn Ross Akad Med Nauk 1998;3:6-9.

179 Karpenko LI, Ignat'ev GM, Kozhina EM, Kashentseva EA, Poryvaeva VA, Melamed NV, Baiborodin SI, Smirnova Olu, Il'ichev AA: Isolation and study of recombinant strains of Salmonella typhimurium SL 7207 producing HBcAg and HBcAg-HBs (in Russian). Vopr Virusol 2000;45:10-14.

180 Stuyver L, De Gendt S, Van Geyt C, Zoulim F, Fried M, Schinazi RF, Rossau R: A new genotype of hepatitis B virus: Complete genome and phylogenetic relatedness. J Gen Virol 2000;81:67-74.

(handwritten annotations) Solutions: ① p. ossic particle ② remove HBc₈₀₋₉₀ (β-strand) ③ space @ C-term of insert (not containing β-strand hydrophobic large aa) GSG DE GG

Amino
Acids
© Springer-Verlag 2000
Printed in Austria

Insertion of foreign epitopes in HBcAg: how to make the chimeric particle assemble

L. I. Karpenko[1], V. A. Ivanisenko[1], I. A. Pika[1], N. A. Chikaev[2], A. M. Eroshkin[1], T. A. Veremeiko[1], and A. A. Ilyichev[1]

[1] Institute of Bioengineering, State Research Center of Virology and Biotechnology "Vector", Koltsovo, Novosibirsk Region, Russia [2] Novosibirsk Institute of Bioorganic Chemistry, Sibirian Division of the Russian Academy of Sciences, Novosibirsk, Russia

Summary. Hepatitis B core antigen is one of the most promising protein carriers of foreign epitopes of various human and animal pathogens. Chimeric HBcAg particles can be used as effective artificial immunogenes. Unfortunately, not all chimeric proteins are able to be particulated. The dependence of correct or incorrect folding of chimeric proteins on physical and chemical properties of inserts was studied with the help of ProAnalyst, SALIX and QSARPro computer programs. We have found that insertion of amino acids with high hydrophobicity, large volume, and high β-strand index prevent self-assembling chimeric proteins. These factors are most important for the C-termini of inserts. Recommendations for obtaining correct folding of chimeric HBcAg particles have been given.

Keywords: Amino acids HBcAg – Foreign epitopes – Self-assemble – Core particles

Introduction

The design of molecular vaccines on the basis of peptides that are epitopes of infection agents is a promising trend in vaccinology. Unfortunately, such peptides exhibit rather low immunogenicity and therefore require coapplication of adjuvant to induce strong immune response. Immunogenicity can be increased via presenting target sequences in several copies on the surfaces of recombinant virus or virus-like particles (Lomonossoff and Johnson, 1995; Ulrich et al., 1998).

Hepatitis B core protein (HBcAg) is one of the most promising delivery vehicles of foreign epitopes suitable for designing highly immunogenic vaccines (Clarke et al., 1987; Pumpens et al., 1995; Schodel et al., 1996). HBcAg consists of identical 21-kDa protein subunits, which are able to spontaneously assemble into a core particle. Recent electron 3D microscopy

330 **L. I. Karpenko et al.**

suggests that there are two kinds of core particle: 34 and 30 nm in diameter, having T = 4 and T = 3 symmetry and containing 240 and 180 protein subunits, respectively (Crowther et al., 1994).

There are several reasons to use the HBc protein as a carrier of foreign epitopes. First, this can be readily availability in large amounts and is able to spontaneously assemble into a perfect core particle, whether expressed in pro- or eucaryotes. HBc-protein self assembly does not require other viral components. Secondly, HBcAg is highly immunogenic, it induces strong B cell, T cell and CTL responses in human and immunized animals. HBcAg may act either as a T-dependent or as a T-independent immunogen, which directly activates B cells (Milich and McLachlan, 1986). Furthermore, chimeric HBcAg particles enhance the immune response to the inserted foreign epitope, which is presumably because the epitope is presented in many copies on the surface of core particles each containing 180–240 HBc subunits (Clarke et al., 1987).

Foreign epitopes were inserted into HBcAg in various protein regions, including the N-, C-termini and the immunodominant e1 loop (Pumpens et al., 1995). It has been demonstrated that the loop in the main determinant of the core antigen is the most promising insertion site from the immunological point of view (Schodel et al., 1991; Karpenko and Ilyichev, 1998). Epitopes inserted there possess higher antigenicity and immunogenicity than anywhere else. Unfortunately, not all of chimeric proteins are able to be particulated. The ability of chimeric HBcAg to self-assemble is therefore most likely to depend on the physical and chemical properties of the amino acid residues forming the inserted foreign peptide. Although the problem is well known, the dependence has not been studied in detail. Herein we report correlation between various properties of amino acids inserted in the e1 loop and the ability of chimeric HBcAg to assemble into virus particles.

Material and methods

Peptides were inserted in the immunodominant e1 loop of HBcAg. For convenience, those not precluding particle assembly are referred to as "positive"; those doing so as "negative"; the chimeric particles that can still assemble as "viable"; throughout.

Analysis of amino acid sequences of the inserts was performed using the programs ProAnalyst (Eroshkin et al., 1995), SALIX and QSARPro (Ivanisenko, 1998), of which the latter two are our recent developments.

ProAnalyst is a software tool for studying the structural and functional organization of proteins and the correlation of structure and activity. Also, it predicts functionally important amino acid substitutions in peptide alignments or protein alignments, calculates a large number of physical and chemical properties of sites in primary and tertiary protein structures, specifies those important with respect to activity and plots the structure/activity dependencies.

QSARPro handles ProAnalyst output data on structure/activity relationships and generates a list of amino acid substitutions deemed optimal for the user to attain the desired modification of protein activity.

ProAnalyst and QSARPro have statistical data processing capabilities thereby enabling multiple regression, discriminant and variance analyses.

SALIX performs multiple structure alignment of protein sequences, calculates structural parameters and outputs physical and chemical profiles and multiple alignment data in a handy format. SALIX performs multiple alignments of primary and secondary structures, physical and chemical profiles.

We used these programs to calculate the correlation between the physical and chemical properties of the inserts and the viability of chimeric particles. Statistical data processing was performed using multiple linear regression, discriminant and variance analyses. Both methods yielded similar results, we will only refer to those of regression analysis throughout. To enable regression analysis, 1 was assigned to the ability to assemble, and 0 to failure to do so.

The study was conducted on a large variety of physical and chemical properties of amino acids, namely hydrophobicity, volume and polarity (Bogardt, 1980), Chou-Fasman parameters (propensity for α-helices, β-strands, β-turns) (Chou and Fasman, 1978), charge and others. Physical and chemical properties of the inserts were calculated by averaging those of their amino acid sequences.

Results

Analysis of relationships between amino acid factors and the viability of chimeric core proteins

All the information on positive and negative inserts available to us was gathered into a database (Table 1).

As can be seen, the lengths of positive inserts vary between 4 and 111 amino acid residues, which suggests that this parameter is not a factor in the dilemma "to fold or not to fold". In the face of it, length was not included in the analysis.

The search for a relationship between insert structure and core particle viability was conducted using a sliding frame of a length ranging from 1 to the full length of the inserts in the course of analysis of the physical and chemical properties of both full-length sequences and their fragments.

Analysis of full-length inserts revealed a correlation of viability with the β-strand index, hydrophobicity and volume of the amino acids in the inserts. Where the values of these insert parameters are high, the chimeric HBcAg particles are normally not viable. The correlation coefficients fall within an interval between 0.4 and 0.53.

The most convincing correlation was revealed for region 1-7 of the C-terminus (see Table 1). From among all the correlates found, the most statistically significant are the β-strand index, hydrophobicity and volume. Not only does the same refers to the full-length sequences, but their respective dependencies behave alike: there higher the values, the lower the viability of chimeric proteins (see Fig. 1).

These data provide further support to the hypothesis about an important role of the hydrophobicity and β-strand index of the insert in the preservation of the native conformation of chimeric proteins (Gren and Pumpen, 1988; Makeeva et al., 1995).

Noteworthy, no statistically significant correlation were found for the N-termini of the inserted peptides. It is likely that the structure of a native protein imposes by far less harsh structural requirements on the N-terminus than on the C-terminus of the inserts.

332   L. I. Karpenko et al.

Table 1. Sequences of inserts into immunodominant loop of HBcAg

No	Name'	Refs.	Sequence of inserted peptide	Viability
*	e1 loop of HBcAg	Argos et al. (1988)	(74)VNLEDPASRDLVVSYVNTN(92)	1
1.	pnsd	Nassal (1988)	VNLEDpnsdPASRDLVVSYVNTN	1
2.	HPV 16 ET	Tindle et al. (1994)	VNLEDPASeynldssRDLVVSYVNTN	1
3.	HBV HBsAg	Karpenko (1998)	VNLEDPISictftpidgacsRDLVVSYVNTN	1
4.	HBsAg preS2	Schodel et al. (1990)	VNLEDPASKdprvrglyfpaDLVVSYVNTN	1
5.	HRV2 VP2	Clarke et al. (1991)	VNLEDPASvkaetrlnpdllqpteeasRDLVVSYVNTN	1
6.	HIV gp41	Clarke et al. (1991)	VNLEDPASdrpegieeegqerdrdrasRDLVVSYVNTN	1
7.	HPV 16 ET	Tindle et al. (1994)	VNLEDPASededidgpaggqepdirahymasRDLVVSYVNTN	1
8.	HPV 16 ET	Tindle et al. (1994)	VNLEDPASeynldgidgpaggqepdrahymasRDLVVSYVNTN	1
9.	HBV preS1	Makeeva et al. (1995)	VNLEDPISrgvdpafrwdprvrglyfpardpisRDLVVSYVNTN	1
10.	HBV preS1	Makeeva et al. (1995)	VNLEDPISrvgypastnrqsgrqpdprvrglyfpardpisRDLVVSYVNTN	1
11.	HBV HBsAg	Borisova et al. (1993)	VNLEDgfpessitstgpqrtcttpaqgtsmypscccttpsidgactcRDLVVSYVNTN	1
12.	SPAG-1	Boulter et al. (1995)	VNLEDPASvktdvseehqvgfsdksdpssrtpnahpaelpeelviqnvpedpskvtpiqpsalp qvpttgepgidgtttgpsgpgngegkdllkegckkegfiqfttmtlbpgfevaa RDLVVSYVNTN	1
13.	FMD(A22)V VP1	Karpenko (1993)	VNLEDPISrthptttqtiapsRDLVVSYVNTN	0
14.	VEE E2	Karpenko (1998)	VNLEDPISindvkvwtnssRDLVVSYVNTN	0
15.	LCMV gp272–293	Schodel (unpubl.)	VNLEDPASRIsdssgvenpgzyckttvmlaDLVVSYVNTN	0
16.	HPV 16 ET	Tindle (unpubl.)	VNLEDPASededidgpaggqepdrahymifdllasRDLVVSYVNTN	0
17.	HBsAg preS1	Schodel et al. (1991)	VndhqldpafgpansnpdwdfnpvkddwpSRDLVVSYVNTN	0
18.	HBV preS2	Makeeva et al. (1995)	VNLEDPISrvgvdprvrglyfpadprvrglyfpardpisRDLVVSYVNTN	0
19.	FMDV(A22)VP1	Nekrasova (unpubl.)	VNLEDtjdpngtgkysagsmgrgdllepfarvaaghwvyhrdPISRDLVVSYVNTN	0
20.	HBV preS1 × 3	Makeeva et al. (1995)	VNLEDPISrvgvpastnrqsgrqppasfnrqsgrqpdprvrglyfpardpis RDLVVSYVNTN	0

#Source of foreign epitope; *1* viable insert; *0* not viable inserts; * the region of HBcAg loop to be inserted. Capital letters indicate HBc sequence; small letters correspond to inserted sequences. C-terminal parts of inserts are underlined. *HBV* hepatitis B virus; *HBcAg* core antigen of HBV; *HBsAg* surface antigen of HBV; *HPV* human papilloma virus; *HIV* human immunodeficiency virus; *VEE* Venezuelan equine encephalomyelitis virus; *FMDV* foot and mouth disease virus; *SPAG* theileria annulata sporozoite antigen; *CMV* cytomegalovirus; *HRV* human rhinovirus.

Volume

Hydrophobicity

β-strand index
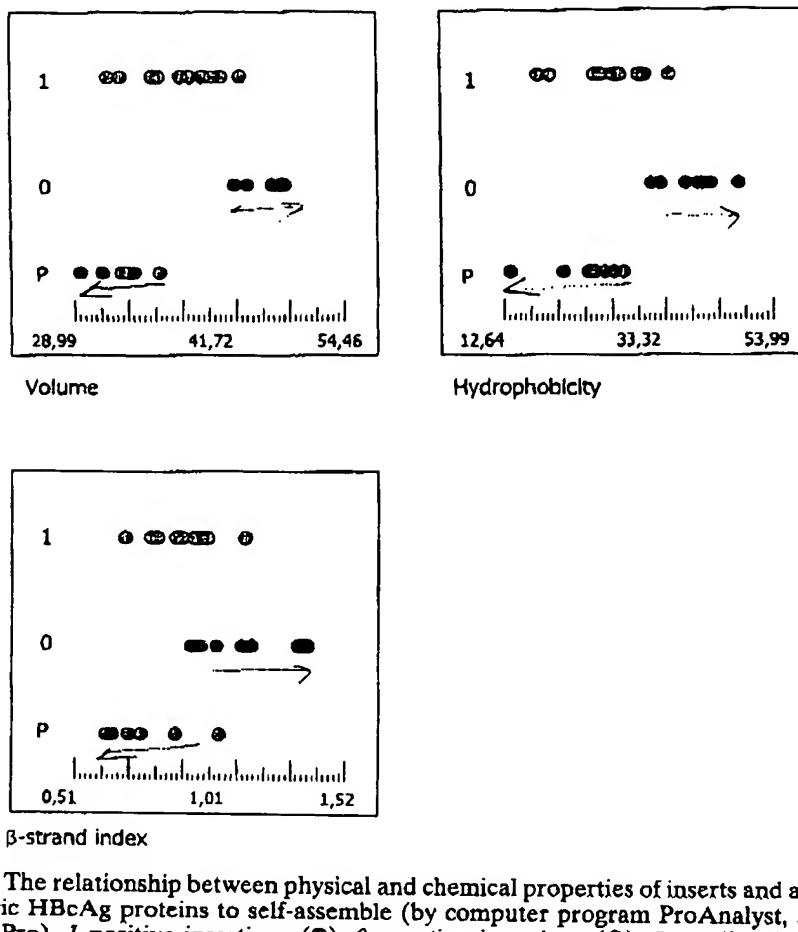
Fig. 1. The relationship between physical and chemical properties of inserts and ability of chimeric HBcAg proteins to self-assemble (by computer program ProAnalyst, SALIX, QSARPro). *1* positive insertions (⊕); *0* negative insertions (●); *P* predicted negative inserts (◑) with GSGDPGG spacer at the C-terminus

Discussion

Analysis of the carrier protein structure
How to make the chimeric particle assemble

There are different ideas of what the tertiary structure of the hepatitis B core protein is like. Argos and Fuller (Argos and Fuller, 1988) describe HBcAg as mostly β-strand protein, whereas Crowther et al. (Crowther et al., 1994) see it mostly α-helical. Despite this difference, the authors agree that spontaneous assembly is preceded by the formation of the HBc dimers. Furthermore, the 76–87 fragment of the main antigen determinant of HBcAg is a loop under either model.

According to the Crowther et al. (Crowther et al., 1994), the loop region, into which peptides were inserted, has high conformational flexibility. The

334   L. I. Karpenko et al.

result of it, we believe, should be a higher probability of the loop being in the contact region of two dimer subunits, provided that the inserts have high hydrophobicity. Obviously, the probability of a perfect dimer is much lower under these circumstances. If the contact region include two loop regions with hydrophobic inserts within, a perfect dimer is yet less probable. The fact that one of the factors influencing the viability of chimeric particles is the hydrophobicity of the inserts does corroborate the hypothesis (Fig. 1).

We propose that the problem of assembly in the situation as described above can be solved by using mosaic particles. Any such particle would consist of two types of dimer, natural and chimeric. Any chimeric dimer, in turn, would consist of different molecules, one native and one chimeric. In this case, a preference for the contact region would only be attributable to the loop region of the subunits carrying the insert, because the loop of the native core protein is hydrophilic.

This hypothesis was verified on a negative insert, which was the Venezuelan equine encephalomyelitis (VEE) E2 epitope. According to our calculations, the VEE epitope has high hydrophobicity, which may account for the inability of chimeric HBcAg-VEE to self-assemble. To produce mosaic capsids, E.coli cells carrying the plasmid that codes for the chimeric HBc-VEE protein were infected with a bacteriophage carrying the native HBcAg gene. Following simultaneous expression of the two genes in one E.coli cell, viable mosaic particles were produced (Loktev et al., 1996).

As was demonstrated above, particle viability also depends on the tendency of inserts to form β-strand. As is known, spatially close linear regions in proteins with high β-strand index can form β-sheets. Naturally, we wanted to see if the HBcAg sequence contains β-strands indeed. As was noted, HBcAg secondary structure is a point at issue (Argos and Fuller, 1988; Crowther et al., 1994). The prediction we did using the PHD program (Rost and Sander, 1994) favours the α-helicity hypothesis (see Fig. 3). As to the location of β-strand regions, one of them (Fig. 2) adjoins loop region 76–85. When the peptide is inserted at position 81, its C-terminus adjoins this region too. If the C-terminus has a high preference for β-strand, there is a possibility of insertion regions and carrier regions forming a β-sheet. We assume that this structural element is capable of preventing assembly. The fact that most of the negative inserts has high β-strand in the C-terminal region provides further support to this idea (Fig. 1).

In our opinion, if the β-strand index is high, there are two solutions to the problem. For instance, β-sheet can be prevented by removing the β-stranded fragment of HBcAg. This was verified by experimenting with a negative insert carrying the epitope of the VP1 protein of the foot and mouth disease virus (FMDV) (see Table 1). Insertion of the FMDV epitope renders the chimeric core protein insoluble and detectable only in inclusion bodies. When region 80–90 of HBcAg (presumably, β-stranded) was removed, the chimeric HBcAg-FMDV protein became soluble and able to self-assembly (Karpenko, 1993).

How to make the chimeric particl assemble   335

```
.........1....,....2....,....3....,....4....,....5....,....6
  I |MDIDPYKEFGATVELLSFLPSDFFPSVRDLLDTASALYREALESPEHCSPHHTALRQAIL|
 II |LLLL........HHHHHHLL.LLL.HHHHHHHHHHHHHHHH..LLLLLL.HHHHHHHHH |
III |......ee.....ebbe...e.b...b.eb.e.b.bb.ee...e.eeeb.ee..bb..bb. |

.........7....,....8....,....9....,....10....,....11....,....12
 |CWGELMTLATWVGGNLEDPISRDLVVSYVNTNMGLKFRQLLWFHISCLTFGRETVIEYLV|
 |HHHHHHHHHHHHH..LLLLLL...EEEEEEE.LLL.HHHHHHHHHHHH.H.HHHHHHHHHH|
 |b..e..e.b....e..eeeee.e.bb..b..e....b..bbbb.bbbb...ee.b.ebb. |
                     ---------
.........13....,....14....,....15....,....16....,....17....,....18
 |SFGVWIRTPPAYRPPNAPILSTLPETTVVRRRGRSPRRRTPSPRRRRSQSPRRRRSQSRE|
 |...EEEELLLLLLLLLLLL.LLLLLL.....LLLLLLLLLLLLLLLLLLLLLLLLLLLLL|
 |.b.b.b...ee...e......e.ee......e...eeee.eee.....e.ee...ee.ee|

.........19
 |SQC|
 |LLL|
 |ee. |
```

Fig. 2. The secondary structure of HBcAg predicted by PHD program. *I* HBcAg amino acids sequence; *II*, secondary structure of HBcAg. Loop signed by *L*; helices by *H*; beta strand by *E*; *III* buried amino acids signed by *b*; exposed ones by *e*

Alternatively, the problem can be solved by using a spacer between the C-terminus of the inserted peptide and the native HBcAg protein. As was mentioned above, it is conditional on such spacers not containing large and hydrophobic amino acids or those with high β-strand index. With these restrictions in mind, we attempted to predict the optimal sequence of a spacer peptide on the basis of the relationships found. The prediction was based on regression equations and classification by Mahalanobis distances, all the procedures were run by QSARPro and ProAnalyst. The spacers selected as eligible by both programs were qualified. In essence, the procedure was a step-by-step filtering of candidate peptides. Because it was the 7 aa C-terminal fragment that proved most significant for spontaneous assembly, the sample comprised random peptides of the indicated length and homologous to the C-terminal fragments of positive inserts. Viability was calculated by regression equations (1–3) for each of these peptides. The candidates were checked for conformity to two requirements: first, viability in excess of 0.8 (which corresponds to viable proteins) following calculations by no matter which equation; second, the physical and chemical properties close to those of the positive inserts.

Eventually, we came up with few peptides, and one of them is GSQDEGG. Our calculations suggest that the fusion of this peptide to the C-termini of negative inserts imparts viability to chimeric proteins (see Fig. 1). Similar results were obtained for other predicted spacer peptides. The one we disclose can be helpful in designing chimeric proteins carrying foreign epitopes.

We hope that our approaches will be helpful to investigators interested in chimeric HBcAg engineering.

336   L. I. Karpenko et al.

Acknowledgements

References

Argos P, Fuller SD (1988) A model for the hepatitis B virus core protein: prediction of antigenic sites and relationship to RNA virus capsid proteins. EMBO J 7: 819–824

Bogardt RA, Jones BN, Durilet FE, Garher WH, Lehman LD, Gurd FRN (1980) Evolution of the amino acid substitution in the mammalian myoglobin gene. J Mol Evol 15: 197–218

Borisova G, Arya B, Dislers A, Borschukova O, Tsibinogin V, Skrastina D, Eldarov MA, Pumpens P, Skryabin KG, Grens E (1993) Hybrid hepatitis B virus nucleocapsid bearing an immunodominant region from hepatitis B virus surface antigen. J Virol 67: 3696–3701

Boulter NR, Glass EJ, Knight PA, Bell-Sakyi L, Brown CGD, Hall R (1995) Theileria annulata sporozoite antigen fused to hepatitis B core antigen used in a vaccination trial. Vaccine 13: 1152–1160

Chou PY, Fasman GD (1978) Prediction of secondary structure of proteins from their amino acid sequence. Adv Enzymol 47: 145–147

Clarke BE, Newton S, Carrol AR, Francis MJ, Appleyard G, Syred A, Highfield PE, Rowlands DJ, Brown F (1987) Improved immunogenicity of a peptide epitope after fusion to hepatitis B core protein. Nature 330: 381–384

Clarke BE, Carrol AR, Brown AL, Jon J, Parry NR, Rud EW, Francis MJ, Rowlands DJ (1991) Expression and immunological analysis of hepatitis-B core fusion particles carrying internal heterologous sequences. In: Brown F, Chanock RM, Ginsberg HS, Lerner RA (eds) Vaccines '91. Cold Spring Harbor Laboratory Press, New York, pp 313–318

Crowther RA, Kiselev NA, Bottcher B, Berriman JA, Borisova GP, Ose V, Pumpen P (1994) Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. Cell 77: 943–950

Eroshkin AM, Fomin VI, Zhilkin PA, Ivanisenko VV, Kondakhin YV (1995) Proanal version 2: multifunctional program for analysis of multiple protein sequence alignments and for stutying the structure-activity relationship in protein families. CABIOS 11: 39–44

Gren E, Pumpen P (1988) Recombinant viral capsids: new generation of immunogenic proteins and vaccines. Zhurnal Vsesojuznogo Khimicheskogo Obschestva Im Mendeleeva (Moscow) 33: 531–536

Ivanisenko VA (1998) Computer analysis of activity and properties relationship in families of similar and mutant proteins. Dissertation, Institute of Molecular Biology, Novosibirsk, Russia, p 35

Karpenko LI (1993) Directed reconstruction of hepatitis B core protein (HBcAg). Dissertation, Institute of Molecular Biology, Novosibirsk, Russia, p 70

Karpenko LI, Ilyichev AA (1998) Chimeric hepatitis B core particles as a presentation system of epitopes of foreign proteins. Vestnik Rus Akadem Nauk (Moskow) 6: 6–9

Loktev VB, Ilychev AA, Eroshkin AM, Karpenko LI, Pokrovsky AG, Pereboev AV, Svyatchenko VA, Ignat'ev GM, Smolina MI, Melamed NV, Lebedeva CD, Sandakhchiev LS (1996) Design of immunogenes as component of a new generation of molecular vaccines. J Biotech 44: 129–137

Lomonosoff GP, Johnson JE (1995) Viral expression systems for peptides. Seminars in Virology 6: 257–267

Makeeva IV, Kalinina TI, Khudyakov YuE, Samoshin VV, Smirnova EA, Semiletov YuA, Pavlyuchenkova RP, Kadoshnikov YuP, Smirnov VD (1995) Heterologous

epitope in the central region of hepatitis B virus core protein. Molekul Biologiya (Moscow) 29: 125–131

Milich DR, McLachlan A (1986) The nucleocapsid of hepatitis b virus is both T-cell independent and T-cell dependent antigen. Science 234: 1398–1401

Nassal M (1988) Total chemical synthesis of a gene for hepatitis B virus core protein and its functional characterization. Gene 66: 279–294

Pumpens P, Borisova GP, Crowther RA, Grens E (1995) Hepatitis B virus core particles as epitope carriers. Intervirol 38: 63–74

Rost B, Sander C (1994) Combining evolutionary information and neural networks to predict protein secondary structure. Proteins 19: 55–72

Schodel F, Thomas W, Will H, Milich D (1990) Recombinant HBV core particles carrying immunodominant B-cell epitopes of the HBV Pre-S2 region. Vaccines '90. Cold Spring Harbor Laboratory Press, Cold Spring Harbor New York, pp 193–198

Schodel F, Will H, Milich D (1991) Hybrid hepatitis B virus core/pre-S particles expressed in live attenuated Salmonella for oral immunization. In: Chanock RM, Ginsberg HS, BrownF, Lerner RA (eds) Vaccines '91. Cold Spring Harbor Laboratory Press, New York, pp 319–325

Schodel F, Kelly S, Tinge S, Hopkins S, Peterson D, Milich D and Curtiss III R (1996) Hybrid hepatitis B virus core antigen as a vaccine carrier moiety. In: Cohen S, Shafferman A (eds) Novel strategies in design and production of vaccines. Plenum Press, New York, pp 15–21

Tindle RW, Herd K, Londono P, Fernando GJP, Chatfield SN, Malcolm K, Dougan G (1994) Chimeric hepatitis B core antigen particles containing B- and Th- epitopes of human papillomavirus type 16 E7 protein induce specific antibody and T-helper responses in immunised mice. Virology 200: 547–557

Ulrich R, Nassal M, Meisel H, Kruger DH (1998) Core particles of hepatitis B virus as carrier for foreign epitopes. Adv Virus Res 50: 141–182

Von Brunn A, Brand M, Reichhuber C, Morys-Wortmann C, Deinhardt F, Schodel F (1993) Principal neutralising domain of HIV-1 is highly immunogenic when expressed on the surface of hepatitis B core particles. Vaccine 11: 817–824

Authors' address: Dr. Larissa I. Karpenko, Institute of Bioengineering SRC Vector, Koltsovo, Novosibirsk Region, 633159 Russia, Fax (3832)-328831, E-mail: karpenko@vector.nsk.su

ClustalW Formatted Alignments

WHV core protein Darrell
Ground squirrel HV

M D I D P Y K E F G S S Y Q L L N F L P L D F F P D L N A L

WHV core protein Darrell
Ground squirrel HV

40 50 60

V D T A . A L Y E E E L T G R E H C S P H H T A I R Q A L V

WHV core protein Darrell
Ground squirrel HV

70 80 90

C W . E L T . L I . W M S   N   T   E . V R   I I V   H V N

WHV core protein Darrell
Ground squirrel HV

100 110 120

T W G L K V R Q . L W F H L S C L T F G Q H T V Q E F L V

WHV core protein Darrell
Ground squirrel HV

130 140 150

S F G V W I R T P A P Y R P P N A P I L S T L P E H T V I R

WHV core protein Darrell
Ground squirrel HV

160 170 180

R R G G . R A . R S P R R R T P S P R R R R S Q S P R R R R

WHV core protein Darrell
Ground squirrel HV

190 200 210

S Q S P . . N C

A molecular assembly system that renders antigens of choice highly repetitive for induction of protective B cell responses

Andrea Jegerlehner[a], Alain Tissot[a], Franziska Lechner[a], Peter Sebbel[a], Iris Erdmann[b],
Thomas Kündig[b], Thomas Bächi[c], Tazio Storni[a], Gary Jennings[a], Paul Pumpens[d],
Wolfgang A. Renner[a], Martin F. Bachmann[a],*

[a] Cytos Biotechnology AG, CH-8952 Schlieren-Zürich, Switzerland
[b] Department of Dermatology, University Hospital, Zürich, Switzerland
[c] Elektronenmikroskopisches Zentrallabor Universität Zürich, Zürich, Switzerland
[d] Research and Study Centre, University of Latvia, Riga LV-1067, Latvia

Abstract

Virus like particles (VLPs) are known to induce potent B cell responses in the absence of adjuvants. Moreover, epitope-specific antibody responses may be induced by VLPs that contain peptides inserted in their immunodominant regions. However, due to steric problems, the size of the peptides capable of being incorporated into VLPs while still permitting capsid assembly, is rather limited. While peptides genetically fused to either the N- or C-terminus of VLPs present fewer assembly problems, the immune responses obtained against such epitopes are often limited, most likely because the epitopes are not optimally exposed. In addition, such particles may be less stable in vivo. Here, we show that peptides and proteins engineered to contain a free cys can be chemically coupled to VLPs formed from the hepatitis B core antigen (HBcAg) containing a lys in the immuno-dominant region. By using this approach steric hindrance of capsid assembly is abrogated. Peptides or protein coupled to VLPs in an oriented fashion are shown to induce strong and protective B cell responses even against self-epitopes in the absence of adjuvants. This molecular assembly system may be used to induce strong B cell responses against most antigens.
© 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Virus-like particles; Molecular assembly system; Protective B cell responses

1. Introduction

Vaccination is one of the most effective means for prevention of infectious diseases. The majority of vaccines used in the past are either attenuated or inactivated forms of the original pathogen, such as the Sabin or Salk vaccine, respectively [1,2]. Most classical viral vaccines, in particular attenuated ones, induce strong cytotoxic T cell (CTL) and B cell responses. The latter response is typically responsible for protection from infection [3]. Although inactivated and attenuated vaccines are relatively safe, there is a small but present risk of reversion to a virulent phenotype in vivo which can cause disease. The most notable examples are observed after vaccination against polio infection. In fact, poliomyelitis is essentially eradicated from the western world and the reported cases are solely vaccine induced [4]. In face of decreasing probability of infection, public acceptance of vaccine-induced side-effects is rapidly declining. Hence, a focus of modern vaccinology has been to produce recombinant vaccines that do not replicate in the host. However, isolated, soluble components of viruses and other pathogens are poorly immunogenic in the absence of non-specific inflammatory stimuli [5] and must be administered together with adjuvant, a notion coined "the immunologist's dirty little secrete" [6]. Since most effective adjuvants are rather toxic or at least painful, only a very limited number have been licensed for use in humans. Thus, there is an urgent need for novel, non-toxic adjuvant formulations.

One reason for the strong B cell responses observed after viral infection is the particular structure of viral surfaces, which are essentially two-dimensional crystals [7,8]. The highly repetitive and organised array of epitopes on viral surfaces efficiently cross-links B cell receptors constituting a strong activation signal that may even overcome B cell tolerance [9–12]. Thus, it may be possible to render an antigen of choice more immunogenic for B cells by presenting them in a repetitive and organised way.

---

* Corresponding author. Tel.: +41-1-733-47-06; fax: +41-61-733-47-07.
E-mail address: bachmann@cytos.com (M.F. Bachmann).

Some viral proteins spontaneously assemble into highly repetitive virus like particles (VLPs) and induce strong B cell responses in the absence of adjuvants [13–15]. Moreover, foreign epitopes inserted into such VLPs induce a similarly strong B cell response, supporting the notion that any antigen in a highly organised array efficiently stimulates antibody responses [16,17]. However, the size and nature of epitopes that can be inserted into VLPs, in particular into their immunodominant regions, is restricted and VLPs containing peptides longer than 20 aminoacids often fail to assemble, beside one remarkable exception, where GFP (238 aa) was inserted into the immunodominant region of HBcAg [19].

In order to address this problem, we have generated a modular molecular assembly system that allows the production of highly organised arrays of various peptides or proteins. The HBcAg which spontaneously assembles into hepatitis B core (HBc), a prototype VLP [13], was modified to contain a lys in the most exposed, immuno-dominant region (position 79/80) and various peptides and proteins were modified to contain a free cys. By chemical cross-linking, these antigens could be directionally placed onto the VLPs. Such antigen-decorated VLPs were observed to induce potent and long-lived immune responses even against self-epitopes in the absence of adjuvants.

2. Materials and methods

2.1. Construction of plasmids ab1 and ab2

Hepatitis B clone pEco63 containing the complete viral genome of Hepatitis B virus was purchased from ATCC. The gene encoding HBcAg was introduced into the *Eco*RI/*Hind*III restriction sites of expression vector pkk223.3 (Amersham Pharmacia Biotechnology AB, Uppsala, Sweden) under the control of a strong tac promotor.

The C-terminus (residues 150–183) of HBcAg which contains the RNA/DNA binding site of the viral capsid was removed to prevent binding of RNA/DNA to the viral capsid. Also a 5 amino acid sequence containing a lys residue was introduced into the c/e1 epitope of HBcAg. The c/e1 epitope (residues 72–88) of HBcAg is located on the surface of the HBc. A part of this region, proline 79 and alanine 80, was replaced by the peptide Gly–Gly–Lys–Gly–Gly. The indroduced lys residue contains a reactive ε amino group in its side chain that facilitates intermolecular chemical cross-linking of HBcAg with any antigen containing a free cys group via a heterobifunctional crosslinker like maleimidobenzoic acid sulfosuccinimidyl ester (Sulfo-MBS). The modified HBcAg was called HBcAg (1–149)-lys. In addition, we mutated both cys 48 and 107 to ser using standard PCR methods. The final plasmid was called ab1.

The M2 peptide (1–23) was genetically fused to the N-terminus of HBcAg (1–183) as described [18]. The final construct was called ab2.

2.2. Expression and purification of HBcAg (1–149)-lys-2cys-Mut and M2 (1–23) HBcAg (1–183)

A culture of *Escherichia coli* K802d containing the plasmid ab1 or ab2 was grown in LB medium (containing 100 μg/ml ampicillin) at 37 °C at 125 rpm until OD_{600} 0.6–0.8 was reached. Expression of the protein was induced by 1 mM IPTG. The cells were solubilised, sonicated, the suspension was centrifuged and the supernatant was precipitated using ammonium chloride. After incubation for 30 min on ice and centrifugation for 15 min at 47,800 × g at 4 °C the supernatant was discarded and the pellet resuspended in PBS, pH 7.2.

The solution containing the HBc was loaded onto a Sephacryl® S-400 HR column (Amersham Pharmacia Biotechnology AB, Uppsala, Sweden), the peak fractions were further purified by CHT® Ceramic Hydroxyapatite column (Bio-Rad, CA, USA). Flow through (which contains purified HBc) was collected. Protein concentration was determined by Bradford assay. From 1 l culture 2–5 mg of purified HBc could be produced.

Electron microscopy was performed according to standard protocols. VLPs were analysed on agarose gels as described [19].

2.3. Expression and purification of GRA2 polypeptide

The gene coding for the 59 c-terminal amino acids of GRA2 with a sequence coding for a C-terminal linker of six amino acids (GSGGCG) were cloned into the pGEX-2T-vector (Amersham Pharmacia Biotechnology AB, Uppsala, Sweden). Expression and purification of the GST-fusion protein was carried out as described in the instructions. GST was cleaved from GRA2 with thrombin while the fusion protein was bound to glutathion–sepharose–beads and the reaction stopped after 20 min with 1 mM PMSF. The sepharose beads were then pelleted by centrifugation and the supernatant containing the GRA2-polypeptide was collected. Protein concentration was determined by Lowry test and concentration of free cys by Ellmann's test. The protein was analysed by SDS-PAGE and westernblot using a monoclonal antibody against the GRA2.

2.4. Expression, purification and refolding of PLA2

The *Pla2* gene containing the H34Q mutation [20] was modified at its C-terminus with a sequence coding for a linker containing a single cysteine residue in its C-terminal part (sequence: AASGGCGG). The modified *Pla2* gene was cloned in the pET11a vector, and transformed into *E. coli* BL21DE3Rill (stratagene). Protein expression and inclusion body preparation and solubilisation were performed as described [21] but for the addition of lysozyme in the cell disruption step, and the use of 200 mM DTT in the solubilisation step. After extensive dialysis against 20 mM Tris, 6 M

Guanidinium–HCl, 0.1 mM DTT, pH 8, the inclusion bodies were reacted with 50 mM oxidised Glutathion for 1 h at RT, and subsequently dialysed against solubilisation buffer lacking DTT. The protein was then refolded by dilution to a final concentration of about 3 µM (estimated by SDS-PAGE) in three portions in a refolding buffer containing Arginine and a redox shuffle containing 5 mM reduced, and 0.5 mM oxidised Glutathion. The protein was dialysed against PBS pH 7.2, concentrated, and purified by gel filtration chromatography on a Superdex G75 column (Pharmacia).

2.5. Coupling procedure

First VLPs were mixed with a 50-fold excess of Sulfo-MBS and incubated for 30 min at RT. Reaction was performed in PBS (pH 7.2). Free, unreacted crosslinker was removed by dialysis with SnakeSkin™ tubing (Pierce, prod. no. 68035) against coupling-buffer (20 mM HEPES, 150 mM NaCl, pH 7.2) at 4 °C overnight. Derivatised VLPs were mixed with a five-fold excess of peptide (100 mM in DMSO) or protein and incubated for 4 h at room temperature. Uncoupled peptide or protein was removed in the same way as unreacted crosslinker by dialysis with a Spectra/Por®6 50 MWCO dialysis membrane (SPECTRUM® LABORATIES INC, CA, USA). Coupling efficiency was determined by SDS-PAGE analysis.

2.6. Immunisation and infection of mice

Six-weeks-old female Balb/c mice were vaccinated with the above antigens in saline without the addition of adjuvants. Mice were immunised on days 0 and 15 with FLAG (CGGDYKDDK) coupled to or mixed with VLPs from HBcAg (1–149)-lys-2cys (100 µg), GRA2 coupled to or mixed with VLPs from HBcAg (1–149)-lys-2cys (50 µg total protein), M2 (SLLTEVETPIRNEWGCRCVGSS-DGGGC) coupled to VLPs from HBcAg (1–149)-lys-2cys or fused to VLPs from HBcAg (1–183) (50 µg), 5′-TNFα-peptide (CSSQNSSDKPVAHVVANHGV) and 3′-TNFα-peptide (SSQNSSDKPVAHVVANHGVGGC) coupled to VLPs from HBcAg (1–149)-lys-2cys (50 µg), or PLA2 coupled to VLPs from HBcAg (1–149)-lys-2cys or PLA2 alone (50 µg). Serum samples were taken at the indicated time points and assessed by ELISA.

Influenza virus PR8 (A/Puerto Rico8/34, H1N1 subtype) was grown, purified and titrated as described [22]. Lung virus titers were determined on MDCK cells as described [22]. Mice were infected intranasally with 10 LD$_{50}$ doses of live virus.

2.7. ELISA

A total of 10 µg/ml peptide coupled to RNAse or 10 µg/ml protein in coating buffer (0.1 M NaHCO pH 9.6) were coated on ELISA plates (Nunc Immuno Maxisorp) and ELISAs were performed according to standard

protocols, using HRPO conjugated secondary antibodies (Sigma). Plates were developed with OPD substrate buffer (0.5 mg/ml OPD, 0.01% H$_2$O$_2$, 0.066 M Na$_2$HPO$_4$, 0.038 M citric acid, pH 5.0; 100 µl each well) and plates were read in an ELISA reader at 450 nm.

3. Results and discussion

3.1. Gene engineering and VLP production

Based upon the predicted structure of HBcAg [23–25] we set about introducing a lys into the most exposed region of HBcAg (position 79/89). This was done to allow chemical crosslinking via the heterobifunctional sulfo-MBS to a free cys introduced into the antigens. The HBcAg was additionally modified by deletion of the aa 150–183 at C-terminus, where the RNA/DNA binding sites of the viral capsid are located [23,26,27] (Fig. 1A). Moreover, internal cys were removed in order to facilitate subsequent assessment of antigen-coupling efficiency (HBcAg (1–149)-lys-2cys). Modified HBcAg (1–149)-lys-2cys as well as wild type HbcAg (1–183) were expressed in *E. coli* and VLPs purified by gel filtration and by hydroxyapatite column. Correct assembly of VLPs was confirmed by agarose gel electrophoresis, where VLP preparations migrated as a single band that stained positive in either EtBr or Commassie Blue staining [19] (Fig. 1B). The absence of an RNA/DNA binding site in the modified version of the HBcAg was confirmed by the absence of EtBr staining (Fig. 1B). Note that the wild type version of the HBc migrates faster than the shorter version since the two particles differ in charge as in the shorter version the arginine rich C-terminus was deleted, a lys was introduced and the internal RNA is missing. Capsid formation was further confirmed by electron-microscopy of negatively stained samples which revealed particles of the expected 34 nm size [24] (Fig. 1C).

3.2. Coupling of a model peptide (FLAG) to VLPs

To assess whether it was possible to couple antigens to the modified VLPs, a model peptide (FLAG), containing a cys at the N-terminus was chemically cross-linked to the modified version of HBc. Densitometric analysis of SDS-PAGE stained with Coomassie Blue demonstrated that 50% of all subunits were coupled (Fig. 2A). Since HBc contains 240 subunits, this efficiency of coupling corresponds to about 120 FLAG-peptides per VLP. By contrast, coupling to the wild type HBc without our lys modification demonstrated <5% coupling efficiency (data not shown). Coupling efficiency of FLAG to VLPs containing internal cys was likely of similar efficiency. However, an accurate estimate of coupling efficiency was more difficult to obtain because of internal cross-linking within the VLPs during derivatisation (not shown).
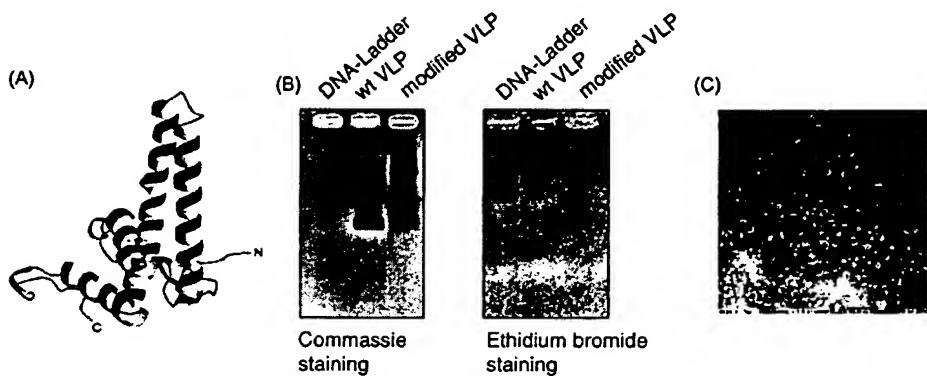
Fig. 1. Assembly of modified HBcAg to HBc. (A) Crystal structure of HBcAg. A lys was placed into the most exposed region of the tip and internal cys removed; (B) modified VLPs exhibiting a lys in the tips of the subunits and lacking internal cys and control wt particles were run in an Agarose gel. The single band observed after Coomassie Blue staining are indicative of correct assembly. Absence of a band in the modified VLP sample after EtBr staining is consistent with deletion of the RNA/DNA binding site. Note that the full length version of the wt HBc runs faster than the short version since the two particles differ in charge as in the shorter version the arginine rich C-terminus was deleted, a lys was introduced and the internal RNA is missing; (C) electron micrograph of assembled particles (modified VLPs).
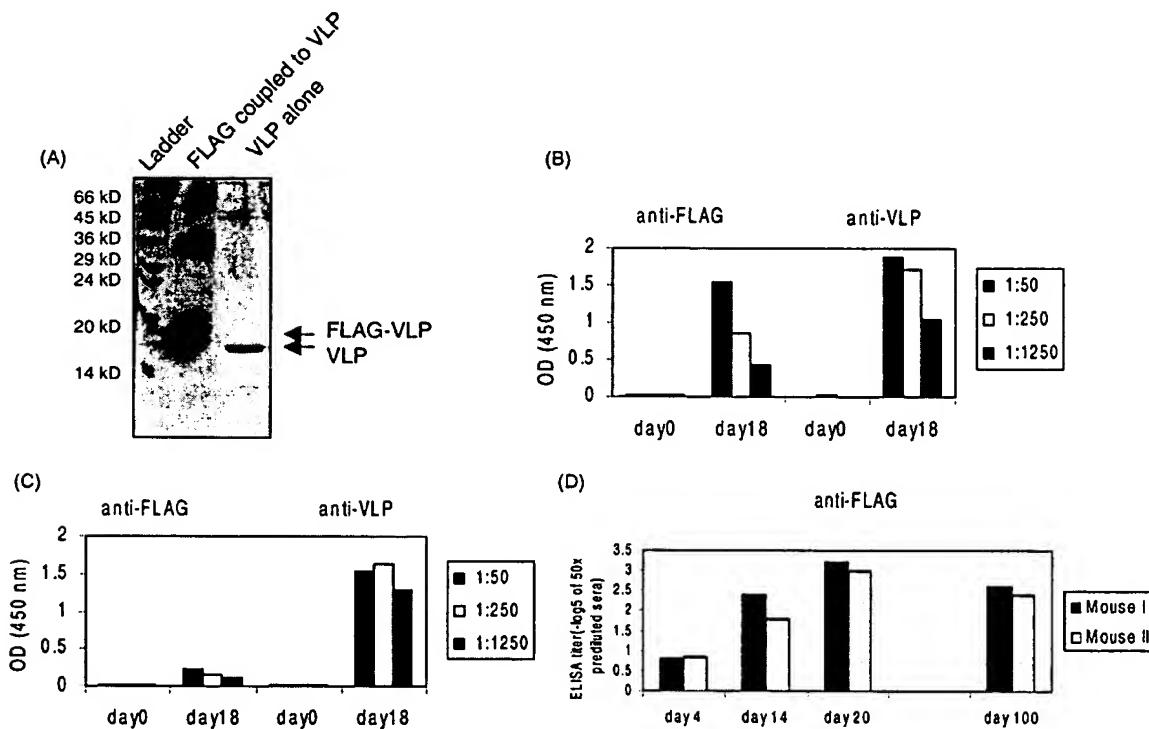
Fig. 2. FLAG-peptide coupled to VLPs induces strong IgG responses. (A) FLAG was coupled to VLPs which were subsequently analysed by SDS-PAGE (12%) under reducing conditions. The relative intensity of the upper band to the lower band is about 1:1 indicating that about 50% of the VLP subunits are chemically coupled to a FLAG-peptide, corresponding to about 120 FLAG-peptides per VLP; (B and C) mice were immunised intravenously with FLAG coupled to VLPs (100 μg total protein) (B) or FLAG mixed with VLPs (C) without adjuvant. Mice were boosted on day 14 with the same amount of antigen. Serum samples were taken on day 0 (pre-immune sera) and day 18. Elisa titers are indicated as dilution at which half-maximal OD was reached. Average titers of three individual mice are shown; (D) mice were immunised intravenously with FLAG coupled to VLPs (100 μg total protein) without adjuvant. Mice were boosted on day 14 with the same amount of antigen and ELISA titers were determined at the indicated time points. Results of two individual mice per group are shown. One representative experiment of four (A–C) or two (D) is shown.

A. Jegerlehner et al. / Vaccine 20 (2002) 3104–3112

3.3. Immunogenicity of FLAG coupled to VLPs

Following coupling, free peptide was removed by dialysis and mice were immunised and boosted intravenously with FLAG coupled VLPs. As a control VLPs were simply mixed with FLAG without chemical crosslinking. Mice immunised with the FLAG-coupled VLPs generated high antibody titers within 18 days of immunisation (Fig. 2B). In contrast, mice immunised with free FLAG mixed with VLPs generated barely detectable immune responses (Fig. 2C). Thus, to obtain high immunogenicity, the antigen should be presented in a highly organised form coupled to VLPs. In fact, the immunogenicity of the FLAG coupled to HBc was similar to the carrier itself, since antibody titers against the VLPs were of a similar order of magnitude to the titers against FLAG (Fig. 2B,C). Titration experiments indicated that 1 μg of peptide-coupled VLPs was sufficient to induce a high IgG titer (not shown). Moreover, antibody responses were long-lived, even though no adjuvants were used for immunisation (Fig. 2D).

3.4. Expression and purification of a small protein (GRA2 domain), coupling to VLP's and study of antibody response

We next assessed whether it was possible to couple larger antigens to the modified VLPs. To this end we cloned a 66 amino acid domain from GRA2, a protein derived from *toxoplasma gondii* [28], as an N-terminal GST-fusion with a C-terminal cys. The recombinant protein was expressed in *E. coli* and purified by affinity-chromatography. The GRA2 domain was protealytically cleaved from GST and coupled to cys free VLPs. Coupling efficiency was assessed by SDS-PAGE. Approximately 30% of the monomers were coupled to GRA2 domain (Fig. 3A). As expected, the coupled VLPs migrated slightly slower than uncoupled VLPs in a non-denaturing agarose gel (Fig. 3B). Mice were immunised and boosted once with the coupled VLPs or with same amount of GRA2 mixed with VLPs. While mice receiving the GRA2 coupled VLPs generated rapid and strong B cell responses, the control mice vaccinated with free GRA2 did not generate any antibody response (Fig. 3C).

3.5. Comparison of protective B cell response of M2 peptide coupled to or fused to VLPs

It has been reported that a peptide derived from the M2 protein of influenza virus is able to induce protective B cell responses in mice [18]. Hence, we investigated whether the M2 peptide coupled to our VLPs was able to elicit a similar response. The M2 peptide, synthesised with a cys at the N-terminus, was coupled to cys free VLPs and used to immunise mice. Coupling efficiency was assessed by SDS-PAGE: approximately 40% of the monomers were coupled to M2 peptide (Fig. 4A). These mice generated a strong IgG response to the peptide (Fig. 4B). The efficiency of the B cell response was compared to the response induced by
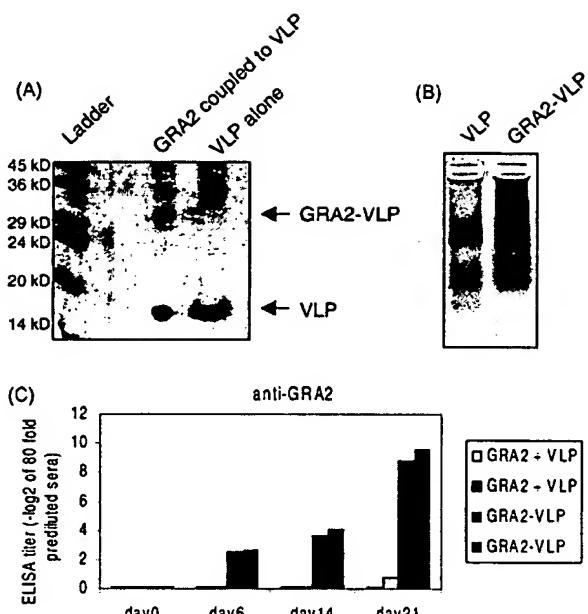
Fig. 3. *Toxoplasma gondii* derived GRA2 domain coupled to VLPs induces a strong IgG response in the absence of adjuvants. (A) A 66 aa domain of GRA2 was purified and coupled to VLPs, and subsequently analysed on a 12% SDS-PAGE Gel under reducing conditions. The relative intensity of the upper band to the lower band was about 3–6 indicating that about 30% of the VLP-subunits are chemically coupled to one GRA2 polypeptide; (B) coupled and uncoupled VLPs were run on agarose gels and visualised using Commassie Blue; (C) mice were immunised intravenously with either GRA2 coupled to VLP (50 μg total protein) or 50 μg GRA2 mixed with VLP and boosted on day14 with the same amount of antigen and ELISA titers were determined at the indicated time points. Elisa titers are indicated as dilution at which half-maximal OD was reached. One representative experiment of two is shown.

M2 peptide genetically fused to the N-terminus of HBcAg (Fig. 4B) [18]. Formation of particles by such modified HBcAg was confirmed by electron microscopy and agarose gel analysis (not shown). With both strategies we mimicked the wild-type structure of M2 protein, in which the free N terminus extends towards the extracellular environment. However, the M2 coupled to HBc was better exposed on the surface of HBc than the fused M2. In contrast, M2 genetically inserted into the immunodominant region would not exhibit the proper structure. The response induced by the coupled M2 peptide was much stronger than that induced by the fused M2 (Fig. 4B). NH₂ terminal sequence analysis by Edmann degradation of the genetically fused M2 construct demonstrated an intact sequence (data not shown). Hence, the weaker immune response generated by the M2-fusion was not a result of proteolytic degradation. Also the difference cannot be explained by the fact that the M2 peptide was fused to the wt HBcAg (1–183), since identical results were obtained when the M2 peptide was coupled to either the short or full-length form of HBcAg (not shown). The lower response induced by genetically fused M2 peptide
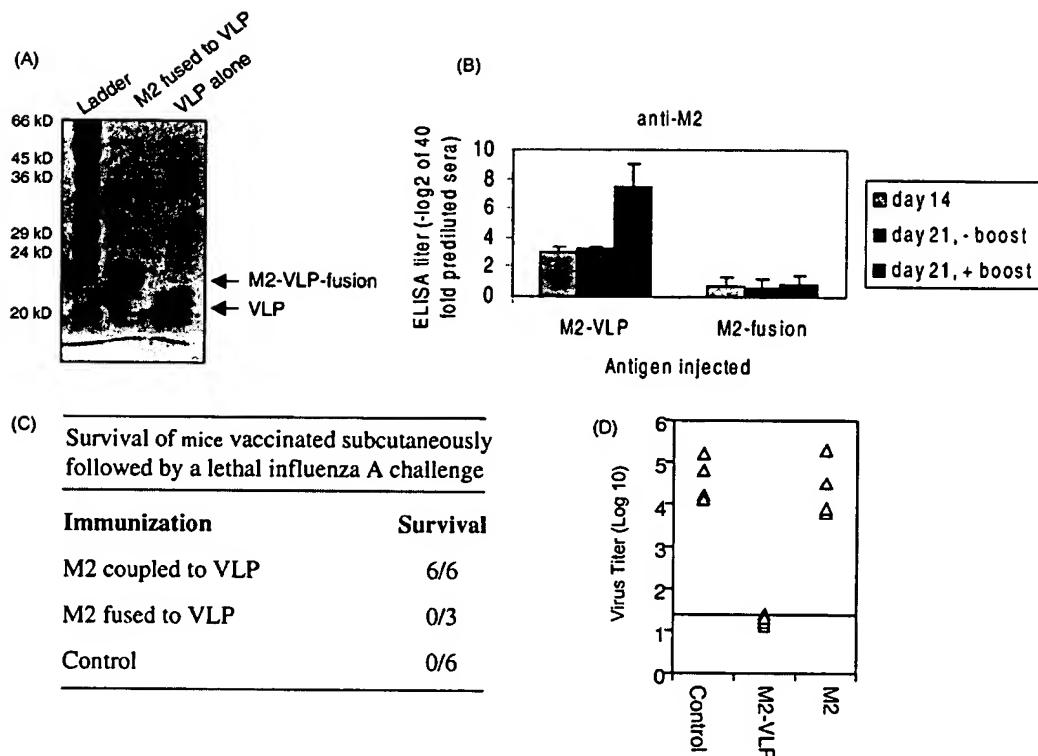
Fig. 4. M2 peptide coupled to VLPs induces a more efficient B cell response than genetically fused M2. (A) M2 peptide was genetically fused to the N-terminus of the full length HBcAg (1–183) and VLPs were assessed on a 12% SDS-PAGE Gel under reducing conditions. Note that HBcAg (1–183) always shows a double band for unknown reasons; (B) eight mice/group were immunised subcutaneously with either M2 coupled to VLP (20 μg total protein) or 20 μg M2 fused to VLP (1–183). Four mice/group were boosted on day 14 with the same amount of antigen. Serum samples were taken on days 14 and 21 after first immunisation for all groups. Average results with standard deviations are shown; (C and D) mice were primed and boosted as in (B) and challenged intranasally 21 days after the boost with life influenza virus. Viral titers were assessed 7 days later in the lung and survival was assessed twice a day over a period of 3 weeks; (D) non-immunized mice and mice immunised with free M2 peptide (10 μg) or M2 fused to VLPs served as controls. Elisa titers are indicated as dilution at which half-maximal OD was reached. One representative experiment of three (B) and two (C and D) is shown.

most likely can be explained by poor accessibility to B cells. Fused M2 is partly buried within the particle while the coupled M2 peptide is maximally exposed on the tip and immunodominant region of the VLP (see also Fig. 1A).

Induction of protective immunity was assessed next. Groups of mice were primed, boosted once with M2 peptide coupled to HBcAg or M2 peptide genetically fused to HBcAg and challenged intranasally 3 weeks later with 10 × LD50 doses of influenza virus. While mice vaccinated with M2 peptide coupled to VLPs were completely protected form lethal challenge with 10 × LD50 doses of influenza virus, mice vaccinated with M2 peptide fused to VLPs were not protected (Fig. 4C). Furthermore, viral titers in the lung of mice vaccinated with M2 coupled to VLPs were very low whereas mice vaccinated either with M2 peptide or control mice exhibited high viral titers in the lung 7 days after challenge with LD50 doses of influenza virus (Fig. 4D).

3.6. Breaking of B cell tolerance by coupling a self-peptide to VLPs

Highly repetitive epitopes have been shown to break B cell tolerance and induce self-specific antibody responses [8–11]. We used a peptide derived from TNFα with either a N- or C-terminal cys coupled to HBcAg to test whether antigens coupled to VLPs in a directed fashion were able to break B cell tolerance (Fig. 5A). The highly organised form of the TNFα-peptide was able to induce a strong antibody response in the absence of adjuvants to the peptide (Fig. 5B,C) that also cross-reacted with native TNFα (Fig. 5D). This indicates that it was possible to induce antibody responses specifically directed against self-molecules. This observation suggests a potential for vaccination with VLPs to block the function of self-molecules, such as TNFα, in order to treat certain chronic diseases including arthritis, colitis and asthma (immunotargeting).
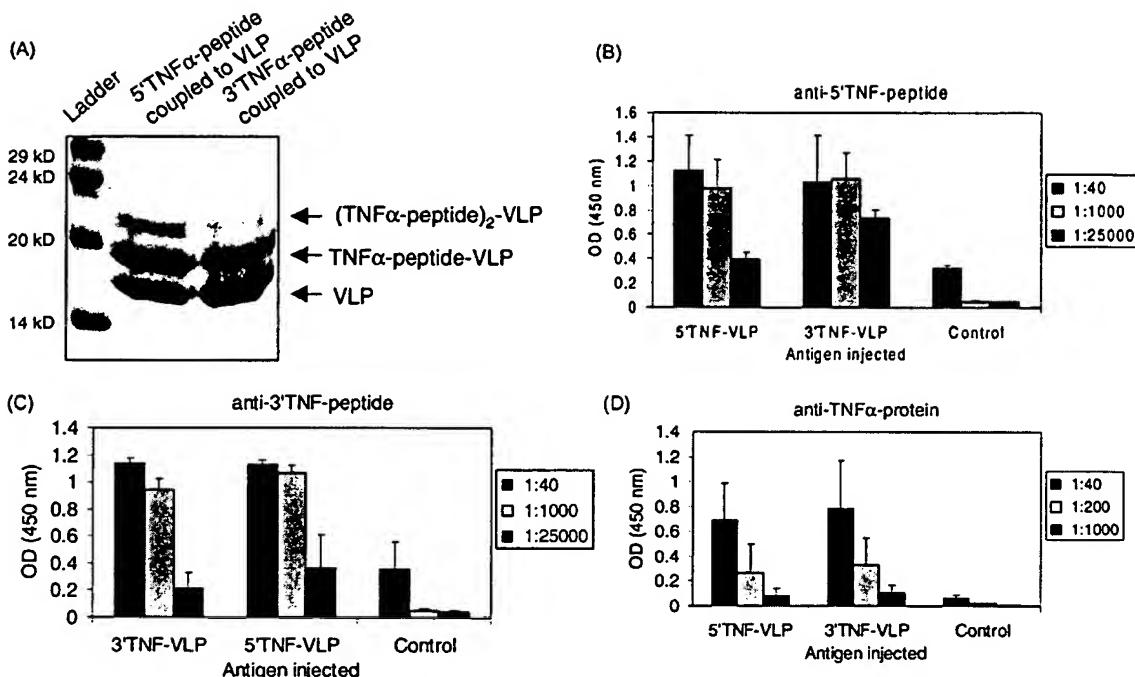
Fig. 5. Induction of self-specific IgG antibodies. (A) TNFα-derived peptides were coupled to VLPs which were subsequently analysed on a 12% SDS-PAGE gel under reducing conditions. The relative intensity of the middle band to the lower band was about 1:1, indicating that about 50% of the VLP-subunits are coupled to a peptide. (B–D) Mice were immunised subcutaneously with either 5′-TNFα-peptide coupled to VLPs (100 µg total protein) or 100 µg 3′-TNFα-peptide coupled to VLPs (100 µg total protein) without adjuvant. Mice were boosted on days 14 and 21 with the same amount of antigen. Serum samples were taken on day 31 after first immunisation and tested for anti-5′-TNFα-peptide IgG-antibodies (B), anti-3′-TNFα-peptide-IgG-antibodies (C) or anti-TNFα-protein-IgG-antibodies (D). Average results of three mice with standard deviation are shown. As a control, pre-immune serum was tested. One representative experiment of two is shown.

3.7. Coupling of full-length proteins and induction of allergen-specific IgG response

Although some protein-derived peptides are able to induce antibodies that recognise the native protein, this is more the exception than the rule. Indeed, with the exception of foot and mouth disease virus [16] and the above mentioned M2-peptide derived from influenza virus [18], virally derived peptides have invariably failed to induce protective B cell responses. Thus, in order to induce pro-
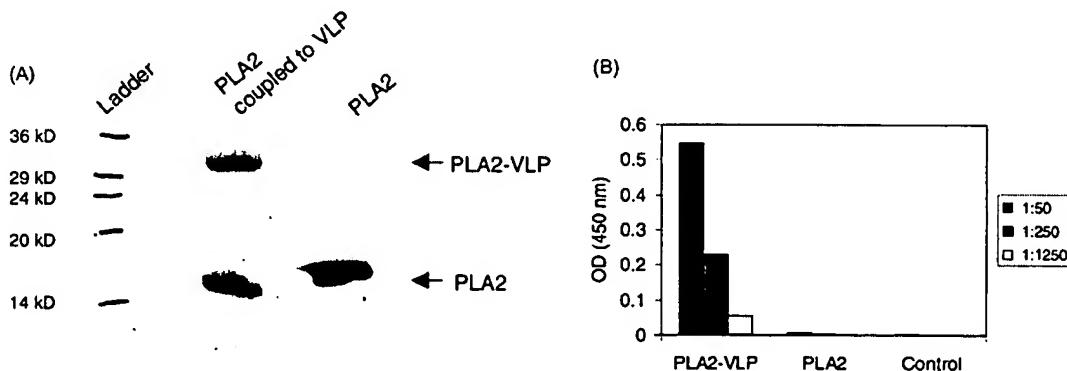
Fig. 6. Induction of IgG-antibodies specific for PLA2, the major allergen in bee venom. (A) Recombinant PLA2 was coupled to VLPs which were subsequently analysed on a 12% SDS-PAGE Gel under reducing conditions of which a western blot specific for PLA2 is shown. In the lane where PLA2 coupled to VLPs was loaded a band representing the coupling product was detected; (B) mice were immunised subcutaneously with 10 µg of VLPs coupled to PLA2 (total protein) of a similar preparation as in (A) or 20 µg PLA2 without adjuvant. Mice were boosted on day 14 with the same amount of antigen and serum samples were taken on day 21. Average results of two mice are shown. As a control, pre-immune serum was tested.

tective immunity, in most instances it is essential to expose the immune system to full length proteins. We, therefore, investigated the coupling of full-length antigens to VLPs. Phospholipase A2 (PLA2) is the major allergene in bee venom and is a biochemically well-defined single-chain glycoprotein consisting of 134 aminoacids. It is primarily responsible for IgE-mediated anaphylactic reaction in allergic patients. It is, however, possible to desensitise patients by inducing a protective IgG response that blocks cross-linking of IgE molecules by the allergens [29]. Hence we tested, whether induction of protective anti-PLA2 IgG responses may be accelerated by coupling the allergen to VLPs. Thus, a catalytically inactive mutant of PLA2 was expressed with a C-terminal cys and expressed in *E. coli*. The protein was expressed in inclusion bodies, solubilised, refolded and purified by size exclusion chromatography. Refolded PLA2 was then coupled to VLPs and used for immunisation of mice in the absence of adjuvants. Analysis of the coupling-reaction by SDS-PAGE was difficult as the band for a HBcAg-Dimer overlapped with the band for PLA2 coupled to HBcAg. However, analysis by western-blot with a monoclonal antibody against PLA2 clarified the situation as a band corresponding to the coupling product could be detected (Fig. 6A). Nevertheless the coupling efficiency was estimated to be rather low. Immunological data supported this estimation as PLA2-specific antibody response of mice immunised with PLA2 coupled to VLPs was rather low compared to the response to the other peptides coupled to VLPs, but still much higher than PLA2-specific antibody response of mice immunised with recombinant PLA2 alone (Fig. 6B). Thus, it is possible that allergen-coupled VLPs may be used for more rapid desensitisation of allergic individuals.

4. Concluding remarks

In summary, a novel technology that allows the coupling of peptides and proteins of choice to VLPs is presented here. Indeed we demonstrated our cross-linking approach to be more effective than the genetic fusion strategy at least in inducing protection against influenza infection. The modular approach that is described permits independent expression of both antigen and carrier, thus, avoiding the problem of steric hindrance of capsid formation observed if the antigen is genetically fused into the VLP. These antigen-decorated VLPs are highly effective vaccines that can induce protective B cell responses against pathogens. Furthermore it is shown that B cell unresponsiveness may be broken and self-specific antibody responses can be induced. Such self-specific B cell responses may allow to selectively target self-molecules that are involved in chronic disease. Finally, allergens coupled to VLPs were able to rapidly trigger the production of allergene-specific IgG responses offering the possibility of accelerated immunotheraphy to allergies. Hence, the decorating of VLPs with target antigens using a modular approach offers novel potential for effective vaccination and immunotherapy.

References

[1] Sabin AB. Oral poliovirus vaccine: history of its development and use and current challenge to eliminate poliomyelitis from the world. J Infect Dis 1985;151(3):420–36.

[2] Salk J, Salk D. Control of influenza and poliomyelitis with killed virus vaccines. Science 1977;195(4281):834–47.

[3] Zinkernagel RM, Bachmann MF, Kundig TM, Oehen S, Pirchet H, Hengartner H. On immunological memory. Annu Rev Immunol 1996;14:333–67.

[4] Murdin AD, Barreto L, Plotkin S. Inactivated poliovirus vaccine: past and present experience. Vaccine 1996;14(8):735–46.

[5] Bachmann MF, Zinkernagel RM, Oxenius A. Immune responses in the absence of costimulation: viruses know the trick. J Immunol 1998;161(11):5791–4.

[6] Janeway Jr CA. Approaching the asymptote? Evolution and revolution in immunology. Cold Spring Harb Symp Quant Biol 1989;54(Pt 1):1–13.

[7] Bachmann MF, Zinkernagel RM. The influence of virus structure on antibody responses and virus serotype formation. Immunol Today 1996;17(12):553–8.

[8] Bachmann MF, Zinkernagel RM. Neutralizing antiviral B cell responses. Annu Rev Immunol 1997;15:235–70.

[9] Bachmann MF, Rohrer UH, Kundig TM, Burki K, Hengartner H, Zinkernagel RM. The influence of antigen organization on B cell responsiveness. Science 1993;262(5138):1448–51.

[10] Chackerian B, Lowy DR, Schiller JT. Induction of autoantibodies to mouse CCR5 with recombinant papillomavirus particles. Proc Natl Acad Sci USA 1999;96(5):2373–8.

[11] Chackerian B, Lowy DR, Schiller JT. Conjugation of a self-antigen to papillomavirus-like particles allows for efficient induction of protective autoantibodies. J Clin Invest 2001;108(3):415–23.

[12] Kouskoff V, Lacaud G, Nemazee D. T cell-independent rescue of B lymphocytes from peripheral immune tolerance. Science 2000;287(5462):2501–3.

[13] Pumpens P, Borisova GP, Crowther RA, Grens E. Hepatitis B virus core particles as epitope carriers. Intervirology 1995;38(1/2):63–74.

[14] Roth JF. The yeast Ty virus-like particles. Yeast 2000;16(9):785–95.

[15] Schiller JT, Lowy DR. Papillomavirus-like particles and HPV vaccine development. Semin Cancer Biol 1996;7(6):373–82.

[16] Clarke BE, Newton SE, Carroll AR, et al. Improved immunogenicity of a peptide epitope after fusion to hepatitis B core protein. Nature 1987;330(6146):381–4.

[17] Fehr T, Skrastina D, Pumpens P, Zinkernagel RM. T cell-independent type I antibody response against B cell epitopes expressed repetitively on recombinant virus particles. Proc Natl Acad Sci USA 1998;95(16):9477–81.

[18] Neirynck S, Deroo T, Saelens X, Vanlandschoot P, Jou WM, Fiers W. A universal influenza A vaccine based on the extracellular domain of the M2 protein. Nat Med 1999;5(10):1157–63.

[19] Kratz PA, Bottcher B, Nassal M. Native display of complete foreign protein domains on the surface of hepatitis B virus capsids. Proc Natl Acad Sci USA 1999;96(5):1915–20.

[20] Dudler T, Machado DC, Kolbe L, et al. A link between catalytic activity, IgE-independent mast cell activation, and allergenicity of bee venom phospholipase A2. J Immunol 1995;155(5):2605–13.

[21] Tissot AC, Pecorari F, Pluckthun A. Characterizing the functionality of recombinant T cell receptors in vitro: a pMHC tetramer based approach. J Immunol Meth 2000;236(1/2):147–65.

[22] Bachmann MF, Ecabert B, Kopf M. Influenza virus: a novel method to assess viral and neutralizing antibody titers in vitro. J Immunol Meth 1999;225(1/2):105–11.

[23] Birnbaum F, Nassal M. Hepatitis B virus nucleocapsid assembly: primary structure requirements in the core protein. J Virol 1990;64(7):3319–30.

[24] Crowther RA, Kiselev NA, Bottcher B, et al. Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. Cell 1994;77(6):943–50.

[25] Wynne SA, Crowther RA, Leslie AG. The crystal structure of the human hepatitis B virus capsid. Mol Cell 1999;3(6):771–80.

[26] Gallina A, Bonelli F, Zentilin L, Rindi G, Muttini M, Milanesi G. A recombinant hepatitis B core antigen polypeptide with the protamine-like domain deleted self-assembles into capsid particles but fails to bind nucleic acids. J Virol 1989;63(11):4645–52.

[27] Hatton T, Zhou S, Standring DN. RNA- and DNA-binding activities in hepatitis B virus capsid protein: a model for their roles in viral replication. J Virol 1992;66(9):5232–41.

[28] Achbarou A, Mercereau-Puijalon O, Sadak A, et al. Differential targeting of dense granule proteins in the parasitophorous vacuole of Toxoplasma gondii. Parasitology 1991;103(Pt 3):321–9.

[29] Lerch E, Muller UR. Long-term protection after stopping venom immunotherapy: results of re-stings in 200 patients. J Allergy Clin Immunol 1998;101(5):606–12.

Quantitative Filter Hybridisation

MARGARET L.M. ANDERSON and BRYAN D. YOUNG

1. INTRODUCTION

An application of nucleic acid hybridisation, which is of central importance to genetic engineering and is finding increasing use in molecular biology, is filter hybridisation. The technique is derived from the classical experiments of Gillespie and Speigelman (1). Denatured DNA or RNA is immobilised on an inert support, for example nitrocellulose, in such a way that self-annealing is prevented, yet bound sequences are available for hybridisation with an added nucleic acid probe. To facilitate analysis, the probe is labelled, often with ^{32}P. Hybridisation is followed by extensive washing of the filter to remove unreacted probe. Detection of hybrids is usually by autoradiography although, when the hybrids are sufficiently radioactive, scintillation counting can be used. The procedure is widely applicable, being used for phage plaque and bacterial colony hybridisation, Southern and Northern blot hybridisation, dot blot hybridisation and hybrid selection (see other chapters in this volume).

While solution hybridisation is the standard method for quantitative measurements of sequence complexity and composition (2), there are practical difficulties when the number of samples is large. By contrast, dot blot hybridisation is ideally suited to the analysis of multiple samples. The technique has the added advantage that it is easy to prepare replicate filters allowing many filter-bound sequences to be analysed at the same time, for example with different probes or under different hybridisation and washing conditions. Dot blot hybridisation can be used qualitatively since it is capable of great discrimination, as exemplified by the ability to distinguish between closely similar members of multigene families (3,4). It can also be used quantitatively with appropriate calibration (5), but it is most commonly used as a semi-quantitative method for determining the relative levels of sequences in different samples. As we shall see, its use is limited by the low rate of hybridisation and by a level of sensitivity which makes it less useful than solution hybridisation, for analysing rare sequences.

In this chapter we will first discuss theoretical aspects of filter hybridisation (Sections 2–5) and then describe practical aspects (Sections 6–14). To make clear the distinction between sequences which are in solution and those which are filter-bound, we will use a different nomenclature from that in the previous chapter. Subscripts 's' and 'f' will be used for nucleic acid in solution and filter-bound nucleic acid, respectively.

2. KINETICS OF FILTER HYBRIDISATION

Nucleic acid hybridisation depends on the random collision of two complementary sequences. As described in the previous chapter, the time course of the reaction in

solution is determined by the concentration of the reacting species and by the second order rate constant, k. The stability of the duplex formed is dependent on its melting temperature, T_m. For hybridisation of perfectly-matched complementary sequences in solution, equations have been derived which describe the process fairly precisely (e.g. Chapter 3, Equations 4, 5 and 10). Detailed investigations have determined the effects of changes in reaction conditions for solution hybridisation, so that values for k and T_m can be calculated with some confidence. In contrast, filter hybridisation has been less extensively studied, and the parameters affecting the rate and extent of reaction are less well understood. Calculations made from solution hybridisation are not necessarily valid for filter hybridisation, although changes in reaction conditions probably have a similar qualitative effect.

The hybridisation of a denatured nucleic acid probe in solution to a filter-bound nucleic acid is a function of two competing reactions, viz. the reassociation of sequences in solution and the hybridisation to filter-bound DNA or RNA. (Since the filter-bound nucleic acid is immobilised, reassociation of bound sequences does not occur.) The rate of disappearance of single strands may be expressed by the equation:

$$\frac{-d[C_s]}{dt} = k_1[C_f][C_s] + k_2[C_s]^2 \qquad \text{Equation 1}$$

where C_f is the concentration of filter-bound nucleic acid sequence, C_s is the concentration of nucleic acid probe in solution. k_1 is the rate constant for the hybridisation reaction on the filter and k_2 is the rate constant for the reassociation in solution. The term $k_1[C_f][C_s]$ represents the filter hybridisation while the term $k_2[C_s]^2$ represents reassociation in solution.

A variety of factors affect the rate constants (see Section 3). The rate constant should be the same for both reactions (i.e....$k_1 = k_2$) provided that:

(i)   the nucleation rates at the filter and in solution are the same
(ii) the effective molecular weight of the nucleic acid species in solution is smaller than that of the filter-bound nucleic acid, since the rate constant for DNA-DNA reassociation is dependent on the size of the smaller fragment (6,7), at least in solution.

Equation 1 predicts that the initial rate of hybridisation is proportional to the concentrations of both the probe in solution and the filter-bound sequences. When $[C_f]$ is much higher than $[C_s]$, as in dot blots of plasmid DNA (this Chapter, Section 6.2.1 and Chapter 5, Section 3.1) or in hybrid selection (Chapter 5, Section 3.2), the solution reassociation term can be ignored and Equation 1 simplifies to the pseudo-first order reaction:

$$\frac{-d[C_s]}{dt} = k_1[C_f][C_s] \qquad \text{Equation 2}$$

where $[C_f]$ is constant.

On integration, this gives:

$$\frac{[C_s]_t}{[C_s]_0} = e^{-k_1[C_f]t} \qquad \text{Equation 3}$$

where $[C_s]_t$ is the value of $[C_s]$ at time t.

While it has been shown experimentally that the initial hybridisation rate is proportional to $[C_s]$, the relationships in Equations 1 and 2 do not describe exactly the dependency of the hybridisation rate on $[C_f]$. At low values of $[C_f]$, the initial rate of hybridisation is proportional to $[C_f]$, but the rate does not increase linearly at higher values (8 – 10). This is explained by the fact that filter hybridisation depends on two processes, diffusion of the probe to the filter and hybridisation at the filter. It is thought that at low values of $[C_f]$, the hybridisation reaction itself is the rate-limiting step, whereas at high values of $[C_f]$ hybridisation is so fast that the solution surrounding the filter becomes depleted of probe and the overall reaction is then limited by diffusion of the probe to the filter. Flavell et al. (9) have shown that at higher values of $[C_f]$ the rate equation should incorporate a term, J, to take diffusion of the probe into account. Therefore, at high values of $[C_f]$, Equation 1 can be replaced by an equation of the form:

$$\frac{-d[C_s]}{dt} = J + k_2[C_s]^2 \qquad \text{Equation 4}$$

where $k_1[C_f][C_s] > J > 0$. The diffusion term J is a function of the diffusion coefficient of the probe and the concentration gradient of the probe. The relationship between J and $k^2[C_s]^2$ determines whether reassociation of the probe is an important factor. When J is $\leq k_2[C_s]^2$, reassociation will be significant.

Since many filter hybridisation experiments aim to hybridise the maximum amount of probe to excess sequences on the filter, the relationship given in Equation 4 is important. The overall hybridisation reaction will be speeded up by factors which increase diffusion of the probe to the filter, for example, using a small probe, high incubation temperature, low reaction volume and shaking the reaction vessel. The rate constant k_1 can be determined by two methods.

(i) From initial reaction rates. Rearranging the terms in Equation 2,

$$k_1 = \frac{v_i}{[C_f][C_s]}$$

where v_i is the initial rate of reaction. So in a filter-bound DNA excess reaction, a plot of the reciprocal of the percentage of added probe which has hybridised to the filter versus the reciprocal of the time of reaction will give a straight line with a slope of $1/k_1$ (Chapter 3, Section 3.3.1). This holds true for both double-stranded and single-stranded probes, although a small correction may have to be made in the value of $[C_s]$ for reassociation of the probe in solution.

(ii) From measuring $t_{1/2}$, that is, the time when $[C_s]_t/[C_s]_0$ is 0.5. For a filter-bound DNA-excess reaction (pseudo-first order kinetics) a plot of log [fraction of the

Quantitative ~~~er Hybridisation

probe remaining single-stranded] against the time of reaction will give a straight line. The $t_{1/2}$ can be read off the graph and substituted in Equation 3 to give:

$$[C_f]t_{1/2} = \frac{\ln 2}{k_1} \qquad \text{Equation 5}$$

or

$$k_1 = \frac{0.693}{[C_f]t_{1/2}} \qquad \text{Equation 6}$$

Experimentally, values for k_1 obtained from $t_{1/2}$ measurements and from initial rate data for hybridisation of a simple DNA probe to filter-bound DNA are in good agreement. However, the values obtained are 10 times lower than those obtained for solution hybridisation of the same DNAs (9). The reason may be that only a fraction of the DNA bound to the filter is accessible for nucleation, although all the DNA can effectively participate in hybrid formation. Thus the concentration term $[C_f]$ used to calculate the rate constant may be incorrect and k_1 (hybridisation) may actually be equal to k_2 (reassociation). Alternatively, rate constants may be lower for filter hybridisation. As a consequence of binding nucleic acid sequences to the filter, steric restraints may retard the formation of stable nucleation complexes.

Equations 1 and 4 show that one of the factors affecting the kinetics of a filter hybridisation reaction is reassociation of the probe. This variable is often overlooked but its effects can be large and can cause problems in interpreting results. It has been shown that as much as 20 – 30% of the input DNA probe can be unavailable for hybridisation due to reassociation (9). A second complication is that the probe may form concatenates of partially-reassociated duplexes with single-stranded regions which can hybridise to filter-bound sequences. Again the effects are not negligible. Flavell *et al.* (11) showed that 10% of the added denatured, double-stranded DNA which hybridised to a filter containing single-stranded DNA represented homologous rather than complementary sequences. Similar problems can arise in DNA-RNA hybridisation experiments with self-complementary transcripts. In order to minimise these complications, it is desirable to choose reaction conditions which facilitate diffusion of probe to the filter and favour hybridisation over reassociation, that is, use of a small probe (preferably single-stranded), small reaction volume, a low concentration of probe in solution and a high reaction temperature.

3. FACTORS AFFECTING THE RATE OF FILTER HYBRIDISATION

3.1 Concentration of the Probe

There has been no systematic study of effects of the concentration of probe on the rate of hybridisation at the filter and on the yield of duplex. However, the following points should be noted.

3.1.1 *Double-stranded Probe in Excess*

If the probe is a simple double-stranded sequence, Equation 1 predicts that at high $[C_f]$ values, reassociation of the probe should be favoured over hybridisation to the bound nucleic acid. Therefore, as incubation continues, the reaction will change from being

in probe excess to being in filter-bound excess where, as we have seen, the kinetics are different. Increasing the concentration of probe in solution, $[C_s]$, will increase the initial rate of hybridisation at the filter and the proportion of filter-bound sequences in duplex will increase, but not dramatically. For DNA probes and filter-bound RNA, as in RNA dot blots, high concentrations of formamide can be used to suppress reassociation in solution (see Section 4.1.2).

3.1.2 *Single-stranded Probe*

Whether in excess or not, there is no reassociation of a single-stranded probe in solution unless there are regions of extensive self-complementarity. The rate of hybridisation to the filter and the amount of hybrid formed should increase with increase in $[C_s]$. It is important to note, however, that the probe concentration should not be increased without limit. If more than about 100 ng ^{32}P-labelled probe per ml is used, non-specific irreversible binding to the filter occurs.

3.2 Probe Complexity

For solution hybridisation, the rate of reassociation of DNA is an inverse function of its complexity, so that the more complex the DNA, the slower the rate of reassociation (2,12). Extending this to filter hybridisations, the rate of reassociation of the probe should fall when the complexity of the DNA increases and its effective $[C_s]$ decreases. This is indeed what is observed (9). In contrast, two effects of complexity are seen for hybridisation of the probe to filter-bound nucleic acid sequences. When $[C_f]$ is low, the rate of hybridisation is inversely proportional to complexity over a 400-fold range, indicating that the reaction is controlled by the nucleation step. However, when the hybridisation reaction is limited by diffusion of the probe to the filter, that is when $[C_f]$ is high, the rate of reaction is independent of complexity (9).

3.3 Molecular Weight of the Probe

For DNA-DNA hybridisation in solution, the rate is directly proportional to the square root of the molecular weight of the nucleic acid (12) and this also describes the relationship of the molecular weight of the probe, but there are insufficient data for an exact relationship to be formulated. The observed rate of hybridisation is significantly depressed by an increase in the molecular weight of a single-stranded probe (which is not capable of reassociation). This effect is even more pronounced when a double-stranded probe is used. This is because the combined effects of a lower rate of hybridisation and the increased rate of reassociation, which accompanies an increase in molecular weight of the sequences in solution, result in lower observed rates of hybridisation and a reduced final yield of hybrid. The difference in dependence on molecular weight of the two types of filter hybridisation is not understood.

3.4 Base Composition

The base composition of nucleic acids affects the rate of hybridisation, the rate increasing with increasing % G+C. However, the effect is small (12) and can be ignored in practice.

3.5 Temperature

The temperature of reaction affects the rate of any hybridisation reaction (13). Typically, a bell-shaped temperature dependence curve is obtained. At 0°C, hybridisation proceeds extremely slowly, but as the temperature is raised, the rate increases dramatically to reach a broad maximum which is 20−25°C below T_m for DNA-DNA annealing. At higher temperatures the duplex molecules tend to dissociate so that as the temperature approaches T_m −5°C, the rate is extremely low. The relationship applies to the formation of both well-matched and poorly-matched hybrids although the curve is displaced towards lower temperatures for mismatched duplexes (14). So, ideally, hybridisation should be carried out at a T_i (incubation temperature) that is 20−25°C below T_m. In practice, for well-matched hybrids, the hybridisation reaction is usually carried out at 68°C in aqueous solution and at 42°C for solutions containing 50% formamide. For poorly-matched hybrids, incubation is generally at 35 − 42°C in formamide-containing solutions.

A similar dependence has been shown for RNA-DNA hybridisations (10), but here the maximal rate of hybridisation is obtained at some 10−15°C below the T_m of the hybrids.

3.6 Formamide

Formamide decreases the T_m of nucleic acid hybrids (see Section 4.1). This is a very useful property because by including 30−50% formamide in the hybridisation solution, the incubation temperature, T_i, can be reduced to 30−42°C. This has several practical advantages: the probe is more stable at lower temperatures, there is better retention of non-covalently-bound nucleic acid on the filter and nitrocellulose filters are less likely to disintegrate at the lower temperature.

Concentrations of formamide between 30 and 50% apparently have no effect on the rate of filter hybridisation and 20% formamide reduces the rate by only about one third (15). On the other hand, a concentration of 80% formamide is thought to depress the rate constant for hybridisation in solution at least by a factor of three for DNA-DNA duplexes and by a factor of 12 for RNA-DNA hybrids (16). Qualitatively similar results are likely to occur in filter hybridisation.

Formamide can be used to alter the stringency of the reaction conditions. By holding T_i constant and varying the concentration of formamide, different effective temperatures are obtained. Effective temperatures as low as 50°C below the T_m of perfectly-matched hybrids can be reached which allows detection of homologies with as much as 35% mismatching (15).

3.7 Ionic Strength

At low ionic strength, nucleic acids hybridise very slowly, but as the ionic strength increases, the reaction rate increases. The effect is most dramatic at low salt concen-

rations (<0.1 M·Na+) where a 2-fold increase in concentration increases the rate 5- to 10-fold. Above 0.1 M Na+ the rate dependence is less, but still marked up to about 1.5 M Na+ (12,17).

High salt concentrations stabilise mismatched duplexes, so to detect cross-hybridising species, the salt concentration of hybridisation and washing solutions must be kept fairly high. Washing is therefore generally carried out using 2−6 x SSC (1 x SSC is 0.15 M NaCl, 0.015 M trisodium citrate, pH 7.0).

3.8 Dextran Sulphate

Wetmur (18) observed that the addition of an inert polymer such as dextran sulphate increased the rate of hybridisation in solution. Thus the presence of 10% dextran sulphate gave rise to a 10-fold increase in reassociation rate. The effect was attributed to the exclusion of the DNA from the volume occupied by the polymer, that is, the dextran sulphate effectively increased the concentration of the DNA. A qualitatively similar effect occurs in filter hybridisation using both DNA and RNA probes (19) where, of course, the concentrating effect of the polymer applies only to the solution phase. For a single-stranded probe, the rate of hybridisation increases by 3- to 4-fold. For a double-stranded probe, the rate apparently increases by up to 100-fold and the yield of hybrid apparently also increases. However, in both cases, most of this increase is caused by the formation of concatenates which readily occurs under these conditions, that is, extensive networks of reassociated probe which, by virtue of single-stranded regions, hybridise to filter-bound nucleic acid and so lead to over-estimation of the extent of hybridisation. For qualitative studies this amplification in the hybridisation signal caused by binding of labelled probe is quite useful. However, for quantitative studies the effect may complicate the interpretation of results. Therefore it may be desirable to reduce the likelihood of networks forming by using probes which are not self-complementary. If double-stranded DNA probes are used they should be short to minimise the formation of extensive networks of probe. For example, if nick-translated, double-stranded DNA probes (Chapter 2, Section 4.1.2) are used, the DNase concentration in the nick-translation reaction should be adjusted to give fragments ≤400 nucleotides long. Short incubation times should also be used since the formation of networks occurs late in the reaction. Finally it should be noted that solutions of dextran sulphate are viscous (and so can be difficult to handle) and can lead to high backgrounds.

3.9 Mismatching

Many hybridisation reactions involve complex mixtures of sequences and the duplexes formed are not all perfectly base-paired. Mismatching has the effect of lowering the rate of hybridisation and the melting temperature of hybrids, T_m. The temperature dependence of k_i, the rate constant for the formation of mismatched hybrids, still gives a bell-shaped curve (Section 3.5), but k_i is lower and reaches its optimum at a lower temperature relative to the rate constant for formation of perfect hybrids (4.14). This has not been studied extensively, but available data suggest that if the reaction is carried out at a temperature which is optimal for the formation of mismatched sequences, that is, about 25°C below their T_m, the rate is reduced by a factor of two for every 10% mismatch (14).

3.10 Viscosity

As the viscosity of the solution increases, the rate of hybridisation decreases. The effect can be quite large. but there is insufficient data to formulate an exact relationship.

3.11 pH

The effect of pH has not been studied extensively but within the pH range 5–9, the rate of hybridisation at 0.4 M Na^+ is essentially independent of pH (ref. 12). In practice. hybridisation experiments are usually carried out at pH 6.8–pH 7.4.

4. FACTORS AFFECTING HYBRID STABILITY

The melting temperature T_m is a measure of the thermal stability of hybrids. No systematic study of the effect of different parameters has been made for filter hybridisation. However, in general. variables that alter the rate constant, k, also alter the T_m in the same direction. The relationships below are derived from studies on hybridisation in solution. but are expected to be similar, qualitatively at least, for filter hybridisation. It is worth noting. however, that as a consequence of binding nucleic acid to the filters. the T_m of hybrids is often lower than would be predicted from solution hybridisation studies (5).

4.1 Perfectly-matched Hybrids

4.1.1 DNA-DNA Hybrids

Many studies on the stability of perfectly-matched DNA duplexes in solution have shown that T_m is dependent on ionic strength. base composition and denaturing agents (14. 20.21). The following relationship has been derived from combining several results (15):

$$T_m = 81.5 + 16.6 \ (\log M) + 0.41 \ (\%G+C) - 0.72 \ (\% \text{ formamide}) \qquad \text{Equation 7}$$

where M is the molarity of the monovalent cation and (% G+C) is the percentage of guanine and cytosine residues in the DNA. The monovalent cation dependence holds between the limits of 0.01–0.4 M NaCl. but only approximately above this (20). The T_m is maximal at 1.0–2.0 M NaCl. The dependence on (% G+C) is valid between 30% and 75% (G+C) (ref. 22). The reduction in T_m by formamide is greater for poly(dA.dT) (0.75°C per 1% formamide) than for poly(dG:dC) (0.5°C per 1% formamide) (ref. 16).

In aqueous solution at 1 M NaCl (equivalent to 6 × SSC), Equation 7 simplifies to:

$$T_m = 81.5 + 0.41 \ (\% \ G+C)$$

The following relationships. derived from solution hybridisation studies, are also useful:

(i) Every 1% mismatching of bases in a DNA duplex reduces the T_m by 1°C (ref. 14).

(ii) $(T_m)_2 - (T_m)_1 \ \mu_1 = 18.5 \log \mu_2/\mu_1$ where μ_1 and μ_2 are the ionic strengths of the two solutions (ref. 23).

4.1.2 RNA-DNA Hybrids

For RNA-DNA hybrids, the term in Equation 7 incorporating formamide concentration does not hold because the relationship between formamide concentration and the depression of T_m is not linear. At 80% formamide, RNA-DNA hybrids are more stable than DNA-DNA hybrids by some 10–30°C depending on the sequence (5,16). Carrying out the reaction in 80% formamide can therefore also be used to suppress formation of DNA-DNA duplexes and preferentially select RNA-DNA hybrids (5,16,24).

4.2 Mismatched Hybrids

The T_m of nucleic acid hybrids is depressed by base mismatching. Values obtained from solution hybridisation studies show that a 1% mismatch reduces the T_m by between 0.5 and 1.4°C (refs. 17.21.22.25.26). The exact figure depends on the (G+C) content of the DNA. The stability of the hybrids also depends on the distribution of mismatched bases in the duplex. Thus if two sequences have 20% base pair mismatch, the hybrid formed between them will have a high T_m if the mismatch is concentrated in one region leaving a long stretch of perfectly-matched duplex. In contrast, the hybrid will be extremely unstable if every fifth base is mismatched.

At high concentrations of salt, mismatched hybrids are more stable than at low concentrations. In practice this is very useful because varying the salt concentration can be used to stabilise or dissociate mismatched hybrids according to the requirements of the experiment.

5. DISCRIMINATION BETWEEN RELATED SEQUENCES

5.1 Stringency of hybridisation

A sizeable fraction of the eukaryotic genome is composed of families of similar, but not identical, sequences. It is often the aim of filter hybridisation studies to distinguish between closely- and distantly-related members of such a family, for example, in screening recombinant libraries or determining gene copy numbers by Southern blots (Chapter 5). In practice this means that reaction conditions must be adjusted to optimise hybridisation of one species and minimise hybridisation of others.

As explained in Section 3.5. bell-shaped curves describe the relationship between the rate of hybridisation and the temperature of incubation for formation of both well-matched and poorly-matched hybrids. For a poorly-matched hybrid, the rate constant is lower and the curve is displaced towards lower temperatures. When the ratio of rate constants (discrimination ratio) for cross-hybridisation and for self-hybridisation is plotted against temperature of reaction, a sigmoidal curve is obtained (Figure 1). At low temperatures, the ratio is high while at higher temperatures (approaching T_m −20°C for perfectly-matched hybrids), the ratio approaches zero (4,14). Although the data are not extensive, Beltz et al. (4) have suggested that this curve is probably a member of a family of sigmoidal curves whose exact dependence on temperature depends on the degree of mismatching of the hybrids. The relationship is useful in that it predicts that it should be easier to distinguish between distantly-related sequences by incubating at low temperatures while it should be easier to distinguish closely-related sequences by hybridising at high temperatures.

depletion of the probe is given by the following equation:

$$-\frac{dC_s}{dt} = k_1[C_f][C_s] + k_2[C_s]^2 + k_i[C_i][C_s]$$

Equation 8

where $[C_s]$ here refers to the concentration of the probe in solution at time zero.

This equation is derived from Equation 1 by the addition of a term to allow for cross-reaction of the probe with a related, filter-bound sequence, i, which has a concentration $[C_i]$ and a hybridisation rate constant k_i. $[C_f]$ is the concentration of filter-bound sequence, f, which is identical to the probe. The kinetics differ considerably depending upon which sequences are in excess and whether the probe can reassociate. For simplicity, in the following analyses (Sections 5.2.1 – 5.2.3) we have assumed that the concentrations of filter-bound sequences are the same (that is $[C_f] = [C_i]$) and $k_1 = k_2$. In fact we know that the rate constants are not equal, but the result will be qualitatively the same.

5.2.1 Filter-bound Nucleic Acid in Excess

When the filter-bound sequences are in excess over the probe, as in typical plasmid DNA dot blots (Section 6.2.1), Equation 8 simplifies to a pseudo-first order reaction where the rate of loss of the probe $(-dC_s/dt)$ is given by:

$$-\frac{dC_s}{dt} = k_1[C_f][C_s] + k_i[C_i][C_s]$$

Equation 9

Since $\qquad [C_f] = [C_i]$

then $\qquad -\frac{dC_s}{dt} = [C_f][C_s] \sum k$

The rate of hybridisation to sequence i equals $[C_f][C_s]k_i$. Therefore,

$$\frac{\text{rate of hybridisation to sequence i}}{\text{rate of hybridisation to all sequences}} = \frac{[C_f][C_s]k_i}{[C_f][C_s]\sum k}$$

Equation 10

$$= \frac{k_i}{\sum k}$$

In a hybridisation reaction, it is more likely that there will be a number (m) of cross-reacting species. The overall reaction can be treated as the sum of a number of independent hybridisations each with a different rate constant and each following pseudo-first order kinetics. When they go to completion, the probe will all be in hybrids. The frac-

Figure 1. Rate of DNA reassociation as a function of temperature. Normal bacteriophage T4 DNA was used to examine the reassociation of perfectly-matched sequences (self-hybridisation: ●—●) and T4 DNA partially deaminated with nitrous acid was used for mismatched sequences (cross-hybridisation: ○—○). The dotted lines are extrapolations assuming that the rates of reassociation are zero at the appropriate T_m for normal T4 DNA under these conditions (0.15 M sodium phosphate buffer) the T_m is 81°C. The discrimination ratio (▲—▲) is the rate constant for cross-hybridisation k_i divided by the rate constant for self-hybridisation, k. Reproduced from reference 4 with permission.

In practice, therefore, to distinguish between the distantly-related members of a family of sequences, hybridisation should take place at a very permissive (relaxed) criterion. To detect closely-related members, the hybridisation should be at a stringent criterion. A single compromise criterion will not be effective because, as we have seen, different members of the family probably have different discrimination *versus* temperature curves. Hybridisation at a relaxed criterion followed by washing under progressively more stringent conditions may be useful for identifying closely-related members of a family, but is not suitable for detecting distantly-related members. This is probably because hybridisation and washing depend on different parameters. Hybridisation depends on the nucleation frequency while washing depends on the thermal stability (T_m) of the hybrids. Thus, a stringent hybridisation followed by a stringent wash is better for detecting closely-related members of a family than permissive hybridisation and a stringent wash.

5.2 Extent of Reaction

In distinguishing between related sequences, it is important to consider the extent of reaction. At first sight, it might appear that the longer the time of incubation the better should be the discrimination, but this is not the case. The following arguments have been made by Beltz *et al.* (4).

When two (or more) filter-bound sequences react with the same probe, the rate of

Quantitative Filter Hybridisation

tion of the probe hybridised to sequence i is given by:

$$\frac{k_i}{\sum_{j=1}^{m} k_j}$$

At any time during the reaction, the ratio of the amounts of probe hybridised to sequence i and to any other filter-bound sequence j is given by the ratio of the rate constants (k_i/k_j) and the ratio is not affected by the time of incubation. Hence the discrimination between related hybrids is not affected by the extent of the reaction because all the filter-bound sequences continuously compete for the same limiting probe (Figure 2).

5.2.2 Single-stranded Probe in Excess

The kinetics of hybridisation are different from that described above when the probe is in excess over the filter-bound sequences, as in typical genomic Southern blots, genomic dot blots, RNA dot blots and Northern blots (Chapters 5 and 6). If the probe is single-stranded and so cannot reassociate (e.g., for M13 or SP6 RNA probes). Equation 8 simplifies to the same form as Equation 9:

$$-\frac{dC_s}{dt} = k_1[C_f][C_s] + k_i[C_i][C_s]$$

If $[C_f] = [C_i]$, then rearranging terms,

$$-\frac{dC_s}{dt} = [C_s] \sum_{i=1}^{i} k_i[C_i]$$

It can be shown that $E_i(t)$, the fraction of filter-bound sequence i actually hybridised at time t, is given by the equation:

$$E_i(t) = 1 - e^{-k_i[C_s]t}$$ Equation 11

and the ratio of the extent of hybridisation of cross-hybridising sequence i, to perfectly-matched sequence f is:

$$\frac{E_i(t)}{E_f(t)} = \frac{1 - e^{-k_i[C_s]t}}{1 - e^{-k_f[C_s]t}}$$ Equation 12

This ratio is not constant but varies with time. The discrimination (that is, the actual extent of cross-hybridisation compared with the hybridisation to perfectly-matched sequences) is maximal very early in the reaction when it equals k_i/k_1, but declines with increasing incubation time as the term at the right hand side of Equation 9 approaches unity (Figure 2). This means that although the homologous reaction is faster and will
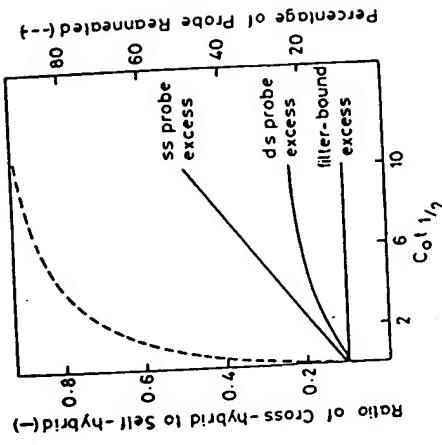
Figure 2. Effective discrimination between perfectly-matched and mismatched sequences as a function of the extent of the reaction. The solid lines represent the ratio $[E_i(t)/E_f(t)]$, see text] of the amounts of probe hybridised to filter-bound mismatched, heterologous sequences (cross-hybridisation) and perfectly-matched homologous sequences (self-hybridisation). Three separate reactions are shown: filter-bound sequences in excess; denatured double-stranded DNA probe in excess; single-stranded probe in excess. The discrimination ratio (k_f/k) is assumed to be the same (0.1) in all cases. The dashed line shows the normal kinetics of reassociation of denatured double-stranded DNA in solution. Reproduced from reference 4 with permission.

reach completion earlier, the heterologous reaction will eventually catch up (single-stranded probe excess: Figure 2). In practical terms, then, with increasing time of reaction, discrimination becomes poorer, so reaction times should be kept short.

5.2.3 Double-stranded Probe in Excess

Consider Equation 8 again:

$$-\frac{dC_s}{dt} = k_1[C_f][C_s] + k_2[C_s]^2 + k_i[C_i][C_s]$$ Equation 8

If the probe can reassociate, the term $k_2[C_s]^2$ is significant. Therefore, as the reaction proceeds, reassociation in solution reduces the amount of probe that is available to hybridise to the filter-bound sequences. So the reaction changes from being in probe excess to one in which filter-bound nucleic acid is in excess. It can be shown that:

$$\frac{\text{fraction of probe hybridising to sequence } i}{\text{fraction of probe hybridising to perfectly-matched sequence } f} = \frac{E_i(t)}{E_f(t)}$$

$$= \frac{1 - (1 + k_2[C_s]t)^{-n}}{1 - (1 + k_2[C_s]t)^{-1}}$$ Equation 13

where $n = k_i/k_1$, that is, the discrimination ratio. Again this means that in practical terms, discrimination equals k_i/k_1 very early in the reaction, but deteriorates rapidly. So, in practice, to distinguish between cross-hybridising species, it is best to use short times of incubation regardless of whether the probe or the filter-bound sequence is in excess. If this does not generate enough signal to be detected, it is advisable to use excess filter-bound sequence and to hybridise for longer times.

6. BINDING OF NUCLEIC ACID TO FILTERS

6.1 Types of Filter Material

There are several types of filter currently in use for the immobilisation of DNA and RNA, for example, nitrocellulose, nylon and chemically-activated papers. The material of choice depends on the purpose of the experiment.

Nitrocellulose filters bind DNA and RNA very efficiently, (~80 $\mu g/cm^2$) except for small fragments of less than about 500 nucleotides in length which are bound rather poorly. The binding procedure is simple. The main disadvantage of nitrocellulose is that it is rather fragile so it requires careful handling and on repeated use tends to become brittle and fall apart. Nylon filters are more pliable than nitrocellulose, are easier to handle and can be used indefinitely without disintegrating. They are reputed to bind nucleic acid as efficiently as nitrocellulose and on the whole we have found that this is true, but we have experienced batch variation in the binding efficiency of some brands. For Southern, Northern and dot blots, both nitrocellulose and nylon filters give excellent results. For DNA dot blots, filters with a pore size of 0.45 μm are used for large nucleic acid molecules and 0.22 μm for molecules of less than 500 nucleotides. For RNA dot blots, filters with 0.1 $-$ 0.22 μm pore size are most efficient.

Although nitrocellulose and nylon filters immobilise nucleic acid, binding by conventional procedures is not covalent. This can lead to problems. For example, nucleic acid is gradually leached off the surface when filters are hybridised for long periods, particularly at high temperature. Furthermore, if the probe in solution is complementary to the entire length of the filter-bound sequence, the hybrid dissociates from the filter and is lost into solution (27). So a consequence of non-covalent binding of nucleic acid is that the hybridisation sensitivity may be reduced with time. New techniques have now been developed for covalent binding of nucleic acid to membranes. This involves u.v. light-induced binding (28). However, to date this is only applicable to nylon filters because of the risk of fire when using nitrocellulose membranes. Chemically-activated paper binds nucleic acid covalently so it has the advantage that it does not discriminate against small nucleic acid molecules and does not lose nucleic acid sequences once they are bound. Both cellulose and nylon filters can be chemically-activated. However, the binding capacity of chemically-activated paper (1 $-$ 2 $\mu g/cm^2$) is much lower than other types of filter and the binding procedure is more complicated. Hence it is not much used for Southern, Northern and dot blots; its main use is in hybrid selection for the enrichment of specific RNA sequences (Chapter 5, Section 3.2).

Suppliers of nitrocellulose filters are Schleicher and Schüll (membrane filters BA85). Millipore U.K. and Waters Associates (Millipore filters), Sartorius Instruments Ltd. (Sartorius filters) and Amersham International plc (Hybond C filters). Suppliers of nylon filters are New England Nuclear (GeneScreen and GeneScreen Plus hybridisation transfer membrane), PALL (Biodyne transfer membrane) and Amersham International plc (Hybond-N membranes). Clearly, this list can never be complete since new products are continually being marketed. The filters are generally available as circles and rectangles in several sizes. Most are also available in rolls which can be cut to size.

All filters require the nucleic acid to be denatured for binding. However, it is most important to note that there is no immobilisation procedure uniformly applicable to all types of filter. Nitrocellulose filters require high ionic strength for quantitative binding of both DNA and RNA and the binding efficiency is much reduced at low ionic strength (29,30). In contrast, GeneScreen nylon membranes require low ionic strength for binding and the binding is poor at high salt concentrations. We have successfully used the same procedures for binding to nitrocellulose and to Biodyne nylon membranes.

6.2 DNA Dot Blots

Multiple samples of genomic or plasmid DNA are spotted next to each other on a single filter in dots of uniform diameter. For quantitative analysis, known amounts of DNA are applied. To evaluate the extent of hybridisation of the probe, a standard consisting of a dilution series of DNA dots is applied in an identical way to the same filter. The procedure binds samples quickly so that many samples can be handled at once. As little as 1 $-$ 3 pg of a hybridising DNA sequence can be detected. Dot blots do not distinguish the number and size of the molecules hybridising, so the hybridisation 'signal' is the sum of all sequences hybridising to the probe under the conditions used.

Commercial apparatus has been developed for binding multiple samples of DNA to filters. A protocol for use of this is described elsewhere in this volume (Chapter 5, Section 3.1). Not every laboratory has access to such a device so the procedure described here involves manual application of samples. This is more time consuming and the dots are less uniform than, when applied by the multiple filtration device, but the results are perfectly satisfactory.

There are many protocols in use for binding samples to filters. They can be divided into two classes according to whether the DNA is denatured before or after it is applied to the filter. Both give satisfactory results. An example of the former method is described for use with the multiple filtration device (Chapter 5, Section 3.1.2). We shall describe an example of the latter method which is in current use in our laboratory with nitrocellulose and Biodyne filters. It is not applicable to GeneScreen filters which require low salt concentration for DNA binding.

6.2.1 Plasmid DNA Dot Blots

It is necessary to convert supercoiled DNA to open circular or linear form to bind to filters. This is because DNA must be single-stranded for binding and denatured super-coiled DNA renatures too quickly on neutralisation to be trapped in the denatured state. Two common ways of obtaining linear or nicked plasmid DNA are to restrict the DNA by enzymic digestion and to treat the plasmid at high temperature (see *Table 1*). The latter partially depurinates the DNA so that on subsequent treatment with alkali the phosphodiester bond breaks at the site of depurination (31). Linear DNA will then separate into single strands.

Nitrocellulose filters are usually treated with high concentrations of salt either at the

Table 1. Linearisation of Plasmid DNA.

Restriction Method

1. Digest the recombinant plasmid with a suitable restriction enzyme. Monitor linearisation of the plasmid by agarose gel electrophoresis.
2. Extract the restricted DNA with an equal volume of phenol pre-saturated with 10 mM Tris-HCl, pH 7.5, 0.1 mM NaCl, 1 mM EDTA. Spin for 2 min in a microcentrifuge to separate the layers.
3. To the aqueous phase, add 0.5 vol. of 7.5 M ammonium acetate and precipitate the DNA by adding 2.5 vol. ethanol pre-cooled to −20°C. Mix well and place at −20°C overnight or −70°C for 1–2 h.
4. Recover the DNA by centrifugation and dry briefly under vacuum.
5. Resuspend the DNA at 50 µg/ml in TE buffer, pH 8.0ᵃ.

High Temperature Method

1. Place 20 µg DNA in a microcentrifuge tube in a final volume of 100 µl 20 mM Tris-HCl, pH 7.4, 1 mM EDTA.
2. Pierce the lid to prevent it popping open and place the tube in a boiling water bath for 10 min.
3. Chill in ice and centrifuge for 10 sec to ensure that all the sample is at the bottom of the tube.
4. Check the volume and adjust to 100 µl with water if necessary so that the DNA concentration remains at 20 µg/100 µl.

ᵃTE buffer is 10 mM Tris-HCl, 1 mM EDTA, pH 8.0.

same time as, or prior to, binding of nucleic acid. This both improves the efficiency of binding and helps to keep the diameter of the dot small. Salts commonly used are 1 M ammonium acetate or 20 x SSC [1 x SSC is 0.15 M NaCl, 0.015 M trisodium citrate, pH 7.0]. A suitable procedure for binding plasmid DNA to nitrocellulose filters involving pre-treatment of the filter with high salt is given below. It is important to note that the filter must not at any stage be handled with bare hands. Grease from the fingers will result in poor binding of nucleic acid and high backgrounds. Therefore disposable plastic gloves must be worn at all stages.

(i) Float a sheet of nitrocellulose on water taking care not to trap air bubbles underneath. When one side is wet, immerse the filter completely to wet the other. If there are dry patches which are reluctant to wet, boil the water for a few minutes.

(ii) Blot the filter lightly on Whatman 3MM paper and transfer to a dish containing 20 x SSC. Leave for 30 min with gentle shaking.

(iii) Dry at room temperature or under a lamp until completely dry.

(iv) If convenient, use a conventional rubber stamp and ink pad to stamp the paper with an array of 5 mm diameter circles to allow easy identification of samples. At this stage, the filters can be stored dry at room temperature sealed in a polythene sleeve.

(v) Place a nitrocellulose filter which has been treated with 20 x SSC onto the lid of a plastic box such that only the edges of the filter are in contact with the lid.

(vi) Apply the plasmid (0.8 µg in 4 µl, linearised as in *Table 1*) to the filter. This can be achieved using a 1–5 µl Supracap pipette (Brand) or an automatic Pipetman (Gilson). Be careful not to puncture the nitrocellulose filter. Keep the diameter of the dots small and do not exceed 4 mm. If necessary, make repeated applications allowing time for each application to dry.

(vii) Allow the samples to dry at room temperature or under a lamp.

Table 2. Reducing the Size of Eukaryotic DNA Prior to Filter Binding.

1. *Either* sonicate 30 µg DNA to an average size of 2 kb or restrict it with a suitable restriction enzyme and then remove the enzyme by phenol extraction (*Table 1*, step 2). The size of the DNA should be checked by agarose gel electrophoresis using appropriate size markers (Appendix II).
2. Recover the DNA by ethanol precipitation (*Table 1*, step 3).
3. Wash the DNA in 70% ethanol, dry briefly under vacuum and resuspend in 1 ml TE buffer, pH 8.0ᵃ.
4. Measure the concentration of DNA spectrophotometrically using the conversion factor $A_{260nm} = 1$ for a solution of 50 µg/ml.
5. Check the volume of solution and freeze dry.
6. Resuspend the DNA in water at a concentration of 10 µg per 4 µl. For serial dilutions, prepare a set of microcentrifuge tubes each containing 6 µl TE buffer, pH 8.0. Remove 6 µl DNA into the first tube containing TE buffer. Mix well. Remove 6 µl from this tube into the second tube with TE buffer and so on until seven dilutions have been made (10 µg–5.3 ng per 4 µl).
7. Centrifuge briefly in a microcentrifuge to ensure that each DNA solution is at the bottom of the tube.
8. Apply the samples to a dry sheet of nitrocellulose pre-treated with 20 x SSC as described in Section 6.2.1, steps (v) – (xii).

ᵃTE buffer is 10 mM Tris-HCl, 1 mM EDTA, pH 8.0.

(viii) Denature the DNA by placing the filter, application side up, on a sheet of Whatman 3MM paper saturated, but not 'swimming', in 1.5 M NaCl, 0.5 M NaOH. Leave for 5 min. (This is conveniently done in a plastic tray.)

(ix) Transfer the filter to Whatman 3MM paper saturated with 0.5 M Tris-HCl, pH 7.4, for 30 sec.

(x) Transfer the filter to Whatman 3MM paper saturated in 1.5 M NaCl, 0.5 M Tris-HCl, pH 7.4, for 5 min. The DNA is now reversibly bound to the filter.

(xi) Place the filter on a dry sheet of 3MM paper and leave to dry at room temperature.

(xii) Sandwich the filter between two sheets of 3MM paper and bake at 80°C for 2–3 h to immobilise the DNA. Ideally, nitrocellulose filters should be baked in a vacuum oven to reduce the risk of fire.

6.2.2 Genomic DNA Blots

Like plasmid DNA, genomic DNA can be applied to filters in dots. The amount of DNA required per dot depends on the experiment being performed. To detect a single copy sequence in eukaryotic DNA, a minimum of 10 µg DNA per dot is suggested. For sequences present in multiple copies, proportionately less can be used.

Modern techniques for DNA isolation usually give a product which is very concentrated and has a very high molecular weight. These two factors make the DNA solution very viscous so that it is difficult to measure the concentration accurately. Hence, to do this, it is necessary to reduce the size of the DNA either by sonication or digestion with an appropriate restriction endonuclease (*Table 2*). This also helps to bind the DNA to the filter more efficiently. Binding of the DNA to nitrocellulose filters, pre-treated with 20 x SSC, is carried out as described above for plasmid DNA (Section 6.2.1).

6.3 RNA Dot Blots

The principle of this procedure is exactly the same as for DNA dot blots. Known amounts of RNA are applied to an inert support and the amount of specific RNA sequence is

Quantitative ... r Hybridisation

determined by hybridisation with a suitable labelled probe. Evaluation of the extent of hybridisation can be made by comparison with standards. The technique is sensitive – as little as 1 pg of a specific RNA sequence can be detected (32). As with DNA dot blots, however, the procedure gives no information on the size or number of sequences contributing to the hybridisation signal. Nylon and nitrocellulose filters are suitable supports. Because nitrocellulose tends to be the most used, its use will be described here. Chemically-activated paper is not generally used for RNA dot blots as its binding capacity is too low.

6.3.1 Preliminary Precautions

One of the main problems of working with RNA is its extreme sensitivity to degradation. Glassware must be scrupulously clean and should never be touched with bare hands which are a good source of ribonuclease. Prior to use, the glassware should be treated with diethylpyrocarbonate to inactivate any ribonuclease. This can be done by immersing the glassware in water to which has been added two drops per litre of diethylpyrocarbonate and then boiling for 30 min. The glassware can then be dried. Heavy metal ions can lead to degradation of RNA especially when long incubations are involved so these should be removed by filtration of all solutions through Chelex resin (BioRad) before use.

6.3.2 Denaturation of RNA and Binding to Filters

Although RNA is single-stranded, it contains double-stranded regions which must be denatured for efficient binding to filters. Alkali treatment is not suitable since it degrades RNA and heat denaturation has not been found to give efficient binding (32). Commonly used denaturants for RNA include glyoxal (33), methyl mercuric hydroxide (34), formaldehyde (35) and dimethyl sulphoxide (DMSO) (36). Since methyl mercuric hydroxide and formaldehyde are toxic and DMSO dissolves nitrocellulose, the procedure described here uses glyoxal as a denaturant. It is based on the methods developed by Thomas (32,37). Glyoxal is supplied commercially as a 40% aqueous solution (6.89 M) which contains polymerisation inhibitors. Glyoxal is readily oxidised to glyoxylic acid which degrades RNA so it is necessary to purify the glyoxal before treating the RNA. This is usually done by deionisation. A suitable protocol is as described in Chapter 6, Table 1.

Glyoxal denatures RNA (and DNA) by binding covalently to guanine residues forming an adduct which is stable at acid and neutral pHs. Glyoxylated nucleic acid binds efficiently to nitrocellulose paper but, after binding, the glyoxal groups must be removed because they have an inhibitory effect on hybridisation. This is easily and quantitatively achieved by treating the filter at 100°C at pH 8.0. Under these conditions as little as 1 pg of a specific sequence of RNA can be detected. The detailed procedure for glyoxalation of RNA and binding to nitrocellulose filters is as follows.

(i) Dry down 20 µg of the sample RNA in a microcentrifuge tube and dissolve it in 5 µl water. The RNA should be salt-free and free of protein which will otherwise react with the glyoxal.

(ii) Prepare a denaturation solution: 34 µl deionised glyoxal, 20 µl 0.1 M sodium phosphate buffer, pH 6.5, 46 µl water.

(iii) Add 5 µl denaturation solution to the RNA. Cover the tube and incubate for 1 h at 50°C. (If using a water bath, make sure that the water comes well up the sides of the tube in order to minimise evaporation.)

(iv) Centrifuge in a microcentrifuge (10 sec) to ensure the sample is at the bottom of the tube. After denaturation in glyoxal, the samples are stable for a few hours and can be kept at 4°C.

(v) If required, make serial dilutions as described in Table 2 (step 6) but using 1% SDS as the diluent.

(vi) Apply the samples by hand to a sheet of nitrocellulose [pre-treated in 20 x SSC as described in Section 6.2.1, steps (i) – (iv)] using the application procedure described in Section 6.2.1, steps (v) and (vi).

(vii) Bake the filter at 80°C for 2 h to immobilise the RNA.

(viii) Remove the glyoxal groups by placing the filter in water at 100°C for 5 – 10 min, then allowing the water to cool to room temperature.

(ix) Blot the filter on 3MM paper and allow to dry at room temperature or under an infra-red lamp.

7. NUCLEIC ACID PROBES

7.1 Types of Probe

In theory, any nucleic acid can be used as a probe provided that it can be labelled with a marker which allows identification and quantitation of the hybrids formed. In practice, double- and single-stranded DNAs, mRNA and RNAs synthesised *in vitro* are all used as probes. Oligonucleotide probes are not used in quantitative dot blots; they are most useful for screening recombinant DNA libraries.

7.1.1 Double-stranded DNA Probes

Double-stranded DNA probes are very commonly used in dot blot analysis. They are often cloned sequences and have low complexity. There are two important points to note when using double-stranded DNA probes:

(i) Two competing reactions occur in filter hybridisation, *viz.* reassociation of the probe in solution and hybridisation to the filter-bound nucleic acid. Therefore, reaction conditions should be chosen to optimise the latter (see Section 11).

(ii) If the DNA is a cloned sequence, it should be excised and purified away from the vector. This avoids complications which can arise if single-stranded vector tails allow formation of concatenates in solution, particularly if the DNA has been randomly sheared. Furthermore, if filters are re-hybridised and the previous probe containing vector sequences has not been completely removed, sandwich hybridisation may occur. That is, duplexes may form through vector sequences rather than through insert sequences. This complicates the interpretation of results.

7.1.2 Single-stranded DNA Probes

With single-stranded DNA probes there is no competing reassociation in solution so single-filter hybridisation is favoured and reactions can be carried out for longer. Single-stranded DNA probes are obtained by strand separation of double-stranded DNA (see Chapter 6, Section 4.2.4) or from M13 phage recombinants.

7.1.3 RNA Probes

RNA probes are more difficult to handle than DNA probes because of the widespread presence of ribonucleases. In addition, mRNA probes, or cDNAs derived from them, are often complex mixtures of sequences and therefore the sequence of interest may represent only a very small proportion of the total nucleic acid. Since the rate of filter-bound hybridisation is inversely proportional to the complexity for low amounts of filter-bound nucleic acid, it may be difficult or impossible to detect the desired hybrids.

Recently it has proved possible to synthesise large amounts of RNA in vitro from specially-constructed recombinant plasmids, such as the SP6 plasmids. The probes have low complexity and, because they are single-stranded, there is no competing reassociation reaction in solution. For these reasons, the use of SP6 RNA transcripts as probes is proving increasingly popular.

7.2 Radiolabelled Probes

Traditionally, filter hybridisations have been carried out with radioactively-labelled probes. ^{32}P is the most commonly used radionuclide and will be the only one discussed here. Conventional labelling replaces a proportion of the nucleotides in a nucleic acid with ^{32}P derivatives or adds ^{32}P to the end of the molecule. After hybridisation, hybrids are detected by autoradiography. ^{32}P has the advantage over other radioisotopes that high specific activities can be readily attained. Much of the technology of filter hybridisation has been developed with it. However, precautions must be taken when handling ^{32}P because of the radiation emitted. Detection by autoradiography, while sensitive, may take a long time if there are few counts in the hybrids. Furthermore, since ^{32}P has a half-life of 14.3 days, experiments should be completed within one half-life.

The preparation of radioactively-labelled DNA and RNA probes, including SP6 RNA transcripts, is described in detail in Chapter 2. In preparing labelled probes for filter hybridisations it is important to remove unincorporated precursors efficiently before use (see Chapter 2) otherwise they may bind non-specifically, but irreversibly, to the filter, giving a high background.

7.3 Non-radioactive Probes

Recent advances in nucleic acid technology now offer alternatives to radioactively-labelled probes. For example, single-stranded DNA can be coupled to a protein. If this protein-DNA complex is now hybridised to filter-bound nucleic acid, the protein in the duplex can be visualised by an antibody reaction (38). If the protein is an enzyme such as peroxidase, then it can be detected and quantitated by its ability to convert a colourless substrate into an insoluble coloured pigment at the site of hybrid formation. This technique is sensitive (1-5 pg nucleic acid can be detected) and has some potentially useful applications. For example, DNA probes coupled to different enzymes can be used in the same hybridisation reaction, so that it should be possible to detect the presence of unrelated sequences simultaneously (38).

Another procedure that uses non-radioactive probes and is becoming increasingly popular is biotin labelling of nucleic acid (39,40). These probes are prepared in a nick-translation reaction by replacing nucleotides with biotinylated derivatives. After hybridisation and washing, detection of hybrids is by a series of cytochemical reactions which finally give a blue colour whose intensity is proportional to the amount of biotin in the hybrid. Biotinylated probes detect target sequences with the same sensitivity as radioactive probes, that is, in the 1-5 pg range. There are several advantages of using biotinylated probes. For example, non-toxic materials are employed and there are no problems of inconveniently short half-lives of the label. This has the additional bonus that biotin-labelled probes can be prepared in advance in bulk and stored at -20°C until required. Detection of hybrids is much faster than for radioactive probes, visualisation of hybrids being complete 2-4 h after washing. One disadvantage of biotin-labelled probes is that the cytochemical visualisation reactions lead to precipitation of insoluble material which cannot be removed, so when the filter is re-used, the previous 'signals' are still present (39,40). The preparation and use of non-radioactive nucleic acid probes is discussed in Chapter 2, Section 4.3.

7.4 Additional Considerations

Additional factors which should be borne in mind when choosing probes are:

(i) It is important to characterise the nucleic acid used for the probe. If any repetitive elements are present (e.g., AluI sequences), they must be removed if the probe is to be used to detect low copy number sequences otherwise hybridisation of the latter will be masked by the repetitive sequence hybridisation.

(ii) The length of the labelled probe is important since the kinetics of hybridisation depend on probe length (see Section 3.3).

(iii) The kinetics of hybridisation differ according to whether the probe or filter-bound sequences are in excess and it is not always immediately apparent which is in excess. What is important is the concentration of the hybridising species, not the total nucleic acid concentration. The following is a rough guideline to this problem. With genomic Southern and Northern blots and genomic DNA and RNA dot blots, the concentration of the probe is likely to be in excess. For example, in Figure 3, even though there is 10 μg RNA per dot and the Ha-ras1 probe is at 20 ng/ml, the probe is in excess for RNA taken from normal tissue. (However, note that the filter-bound sequences are in excess for RNA taken from diseased tissue.) With plasmid or phage dot blots, and phage and colony screening, the filter-bound sequences are likely to be in excess. However, to check which is in excess, the following preliminary experiments can be performed.

(a) Vary the input of probe; if the filter-bound sequence is in excess, the amount of hybridisation should be proportional to the probe input.

(b) Vary the amount of nucleic acid on the filter; if it is in excess, there should be no difference in the amount of probe hybridised.

8. HYBRIDISATION USING RADIOACTIVE PROBES

8.1 Choice of Reaction Conditions

There are many protocols available for hybridising a probe in solution to nucleic acid immobilised on filters. The conditions used depend on the purpose of the experiment and in general are governed by whether DNA-DNA or DNA-RNA hybridisation is involved and whether closely-related or distantly-related sequences are reacting. Reaction conditions that permit formation of hybrids which have a high degree of mismat-

homopolymer tailing. For hybridisations involving RNA, yeast tRNA is often used as a competitor. For *hybridisation*, it is necessary to ensure that the added nucleic acid probe is single-stranded. For double-stranded DNA probes this is usually achieved by boiling or by denaturing in alkali. Radioactive probes can be denatured using either method; heat denaturation is described here and alkaline denaturation is described in Chapter 5 (*Table 3*, step 6). For most purposes hybridisation can be carried out in either aqueous solution or in the presence of formamide. We use the same formamide-containing solutions for both RNA-DNA and DNA-DNA hybridisations, but aqueous solutions for DNA-DNA hybridisations only. Both protocols can be used with nitrocellulose and nylon filters and are described below. After hybridisation, *washing* is carried out to remove unhybridised probe and to dissociate unstable hybrids. The temperature and salt concentration of the washing solution determine which hybrids will be dissociated. In general, washing should be under as stringent conditions as possible; at 5−20°C below T_m for well-matched hybrids and 12−20°C below T_m for cross-hybridising species. In practice, 65−70°C is usually chosen for hybrids having a high degree of homology and 50−60°C for poorly-matched hybrids.

Where possible, the pre-hybridisation, hybridisation and washing steps should be carried out in a shaking water bath or on a shaking platform in an incubator. In filter-bound nucleic acid excess, diffusion of the probe to the filter can be limiting in the absence of agitation. Also, high backgrounds are sometimes encountered if there is no shaking. Solutions should be pre-warmed to the required temperature prior to use.

8.2 Hybridisation in the Presence of Formamide

8.2.1 Pre-hybridisation

(i) To wet the filters evenly, float them on a solution of 1% Triton X-100, taking care to prevent air bubbles being trapped underneath. When one side is wet, immerse the filter to wet the other side.

(ii) Remove the filters and blot gently on Whatman 3MM paper to remove excess liquid.

(iii) It is convenient to carry out the pre-hybridisation and hybridisation reactions in the same container. Typically this is a polythene bag. Suitable bags are 'Sears Seal-N-Save Boilable Cooking Pouches' or Layflat polythene tubing (Trans-Atlantic Supplies). Place each wet filter in a separate bag and heat-seal this, except for one corner, using a domestic bag sealer.

(iv) Add the pre-hybridisation solution (0.08 ml/cm² of filter) which has been prepared as described in *Table 3* and pre-warmed to 42°C. Gently squeeze out the air bubbles and heat seal the corner.

(v) Incubate the filter in the bag for 4−24 h at 42°C. This can be done by placing the bag in a box of water at 42°C in a shaking water bath at the same temperature. Set the water bath to shake at a speed such that the liquid in the bag sweeps gently over the surface of the filter. Alternatively, place the bag on a shaking platform in an incubator at 42°C.

(vi) Cut a corner of the bag and drain the liquid out. Roll a pipette over the surface of the bag to remove as much of the liquid as possible. However, it is most

Figure 3. RNA dot blots. Replicate filters containing 10 μg per dot of poly(A)+ RNA from normal and tumour breast tissue of three different patients were hybridised with the following denatured 32P-labelled double-stranded probes: pT24C3 (Ha-ras-1), pL335 (c-sis) and pHR28 (rDNA). Equal amounts of RNA were present in the dots as judged by the intensities of the hybridisation signals using the rDNA probe. Transcripts homologous to the Ha-ras-1 probe are more abundant in tumour tissue compared with normal breast tissue and this difference is specific to the ras oncogene since the c-sis probe did not give a signal. (Data of Agnatis and Spandidos, with permission.)

ching are said to be permissive (relaxed or low stringency), while those which allow only well-matched hybrids to form are said to be stringent (high stringency). However, the same basic procedure is followed irrespective of the particular reaction conditions used. We will first describe standard hybridisation protocols which have widespread applicability (Sections 8−10) and then how varying the reaction conditions can be exploited to detect different hybrids (Section 11).

The hybridisation process can be divided into three steps: pre-hybridisation, hybridisation with a labelled single-stranded probe, and washing. In the *pre-hybridisation* step, the filter is incubated in a solution which is designed to pre-coat all the sites on it which would bind the probe non-specifically. Failure to do this leads to high backgrounds. Typically, the solution contains Ficoll, polyvinyl pyrollidone and bovine serum albumin [i.e. Denhardt's solution (ref. 41)], and heterologous DNA. As an alternative, heparin can substitute for Denhardt's solution (42). To reduce backgrounds even further, poly(A) and poly(C) are often included. Poly(A) is useful when the probe or filter-bound sequences are rich in A and T residues, e.g., poly(A)+ mRNA or cDNA derived from it. Similarly, poly(C) is included if the probe or filter-bound sequences are rich in G and C residues as when a recombinant is generated through oligo(dG) and oligo(dC)

Table 3. Preparation of Pre-hybridisation Buffer Containing Formamide.

Solution A

Mix together:

Deionised formamide[a]	50 ml
20 x SSC[b]	25 ml
100 x Denhardt's solution[c]	5 ml
1 M sodium phosphate buffer, pH 6.8[d]	5 ml
20% SDS[e]	0.5 ml

Adjust the volume to 95.5 ml with water.

Solution B

Mix together:

Sonicated calf thymus DNA or salmon sperm DNA at 5 mg/ml[f]	2 ml
Poly(C) [5 mg/ml]g	0.2 ml
Poly(A) [5 mg/ml]g	0.2 ml
Yeast tRNA (5 mg/ml)g	2 ml

Denature in a boiling water bath for 5 min.
Quench in ice.

Add solution B to solution A and store at 4°C.

[a]Formamide is a teratogen. Handle with care and use gloves. All contaminated glassware should be soaked overnight in dilute H_2SO_4, then rinsed with water before washing as usual. To deionise formamide, add 200 ml formamide to ~10 g of AG501-X8(D) mixed-bed resin (Bio-Rad). Stir for 1 h at room temperature. Filter through Whatman No. 1 filter paper to remove the resin. Store at 4°C in a dark bottle.
[b]The composition of SSC is 0.15 M NaCl, 0.015 M trisodium citrate, pH 7.0.
[c]100 x Denhardt's solution contains 2% Ficoll (mol. wt. 400 000), 2% polyvinyl pyrrolidone (mol. wt 400 000) and 2% bovine serum albumin. Store at -20°C.
[d]1 M sodium phosphate buffer, pH 6.8, is made by mixing 25.5 ml of 1 M NaH_2PO_4 and 24.5 ml of 1 M Na_2HPO_4. Store at room temperature.
[e]Store this stock solution at room temperature.
[f]Add the DNA to water at ~5 mg/ml. Stir; it may take several hours for the DNA to dissolve. Then sonicate to a length of 400-800 bp. The size can be checked by agarose gel electrophoresis. Adjust the concentration to 5 mg/ml [A_{260nm} = 1 for a solution of 50 μg/ml]. Store at -20°C.
[g]These solutions are stored at -20°C. Their addition to the pre-hybridisation buffer is optional (see Section 8.1). However, there is no disadvantage in adding them even if T- and G-rich sequences are not present in the filter-bound nucleic acid sequence.

8.2.2 Hybridisation

(i) Except for single-stranded probes such as RNA and M13 probes, denature the labelled probe by placing it in a boiling water bath for 5 min. Quench in ice.

(ii) The hybridisation can be carried out in the presence of dextran sulphate which increases the rate of hybridisation (Section 3.8) or in its absence. For hybridisation buffer containing dextran sulphate (prepared as described in *Table 4*), pre-warm the buffer and add the denatured probe to a concentration which does not exceed 10 ng probe/ml or high backgrounds may ensue. In the absence of dextran sulphate, the probe concentration can be increased to 50-100 ng (43). For radioactive probes which have been labelled to a specific activity of

Table 4. Preparation of Standard Hybridisation Solution Containing Formamide.

Solution A

Mix together:

Deionised formamide[a]	50 ml
20 x SSC[b]	25 ml
100 x Denhardt's solution[c]	1 ml
1 M sodium phosphate buffer, pH 6.8[d]	2 ml
20% SDS[e]	1 ml
Dextran sulphate (mol. wt. 500 000)[h]	10 g

Stir until the dextran sulphate has dissolved. Adjust the volume to 95.5 ml.

Solution B

Mix together.

Sonicated DNA (5 mg/ml)[f]	2 ml
Poly(C) [5 mg/ml]g	0.2 ml
Poly(A) [5 mg/ml]g	0.2 ml
Yeast tRNA (5 mg/ml)g	2 ml

Denature in a boiling water bath for 5 min.
Quench in ice.

Add solution B to solution A and store at 4°C.

[a-g]See corresponding footnotes to *Table 3*.
[h]The inclusion of dextran sulphate is optional (see text, Section 3.8).

$1-2 \times 10^8$ c.p.m./μg, a probe concentration of 10 ng/ml gives a solution of about $1-2 \times 10^6$ c.p.m./ml).

(iii) Immediately add this solution to the filter (0.05 ml/cm² filter) and reseal the bag.

(iv) Hybridise the filter at 42°C for the required time. This is normally between 6 and 48 h. Overnight is convenient and for many purposes is sufficient, but see Section 11.2.

8.2.3 Washing

(i) Cut one corner from the bag and remove the hybridisation solution. Retain the probe if it is to be re-used (see Section 13), otherwise discard it down a designated sink.

(ii) Cut open the bag completely and immerse the filter in 200 ml of 2 x SSC, 0.1% SDS at room temperature. Shake the filter gently. Rinse the filter twice for 5 min each time in this solution.

(iii) For a moderately stringent wash, wash the filter twice in 400 ml of 2 x SSC, 0.1% SDS at 60°C for 1 h. For a higher stringency wash, treat the filter for 2 x 1 h at 65°C in 0.1 x SSC, 0.1% SDS.

(iv) Finally, rinse the filter in 2 x SSC at room temperature. Blot the filter to remove excess liquid but do *not* dry the filter if it is to be re-washed or re-screened (Section 13).

(v) Detect the hybrids as described in Section 10.1.

8.2.4 Washing with Nuclease Treatment

In principle, blots can be treated with nucleases to remove unpaired loops and single-

stranded probe tails. This practice should increase the specificity of the reaction, but it has not been studied systematically and there is no real evidence that the treatment is effective. It is better to control specificity by careful choice of reaction and dissociation conditions than through enzyme digestion. DNase and nuclease S1 treatments are not generally used in filter hybridisations, but RNase treatment occasionally is. Filters are treated with a mixture of RNase A and T1 RNase at 25 μg/ml and 10 units/ml, respectively, in 2 x SSC at 37°C for 2 h, then washed in 2 x SSC, 0.5% SDS at 68°C, and finally in 2 x SSC at room temperature.

8.3 Hybridisation in Aqueous Solution

8.3.1 *Pre-hybridisation*

(i)   Wet the filter in 1% Triton X-100 and blot to remove excess liquid.

(ii) Immerse in 4 x SET buffer for 15 min at room temperature. The composition of 4 x SET buffer is 0.6 M NaCl, 1 mM EDTA, 80 mM Tris-HCl, pH 7.8.

(iii) Transfer the wet filter to a plastic bag and heat seal this except for one corner (see Section 8.2.1). Add pre-hybridisation buffer (0.08 ml/cm² of filter), prepared as described in *Table 5* and pre-warmed to 68°C. Incubate at 68°C for between 2 and 16 h.

(iv) Open the bag and remove the pre-hybridisation buffer (see Section 8.2.1 for methodology).

8.3.2 *Hybridisation*

(i)   Denature the probe as described in Section 8.2.2 and add it to fresh pre-hybridisation buffer (pre-warmed to 68°C) at 10—25 ng/ml. Add this solution to the filter at 0.05 ml/ cm² of filter.

(ii) Incubate at 68°C for between 5 and 16 h.

Table 5. Preparation of Aqueous Pre-hybridisation Buffer.

Solution 1

Mix together:

20 x SET buffer[a]	20 ml
100 x Denhardt's solution[b]	10 ml
20% SDS[b]	0.5 ml
5% sodium pyrophosphate	0.1 ml

Adjust the volume to 97.5 ml with water.

Solution 2

Mix together:

Sonicated DNA (5 mg/ml)[b]	2 ml
Poly(C) (5 mg/ml)[b]	0.2 ml
Poly(A) (5 mg/ml)[b]	0.2 ml

Denature in a boiling water bath for 5 min.
Quench in ice.

Add to solution 2 to solution 1 and store at 4°C.

[a] x SET = 0.15 M NaCl, 20 mM Tris-HCl, pH 7.8, 1 mM EDTA.
[b] See relevant footnote to *Table 3*.

8.3.3 *Washing*

(i)   Cut one corner from the bag and remove the hybridisation solution. Retain the probe if it is to be re-used (see Section 13), otherwise discard it down a designated sink.

(ii) Cut open the bag completely. Remove the filter and immerse it in 4 x SET buffer, 0.1% SDS, 0.1% sodium pyrophosphate for 5 min at room temperature.

(iii) Wash the filter three times in 2 x SET, 0.1% SDS, 0.1% sodium pyrophosphate for 20 min each wash at 68°C.

(iv) For a moderately stringent wash, wash the filter three times in 1 x SET, 0.1% SDS, 0.1% sodium pyrophosphate for 20 min each wash at 68°C. For a higher stringency wash, replace the 2 x SET by 0.1 x SET.

(v)   Rinse the filter for 5 min in 2 x SET, at room temperature.

(vi) Blot the filter to remove excess liquid but do *not* dry the filter if it is to be re-washed or re-screened (Section 13).

(vii) Detect the hybrids as described in Section 10.1.

9. HYBRIDISATION USING BIOTIN-LABELLED PROBES

The hybridisation procedure for biotin-labelled probes is essentially the same as for radioactively-labelled probes (Section 8) except that the following points should be noted (39,40):

(i)   The probe should be denatured at high temperature and not with alkali because the amide bond in the linker molecule between the biotin and nucleic acid is alkali-labile.

(ii) Hybridisation is carried out in solutions containing formamide rather than at high temperature. The thermal stability of biotin-labelled hybrids is slightly lower than that of radioactive hybrids. So, in practice, the formamide concentration is lowered from 50% to 45% in otherwise standard hybridisation conditions.

(iii) Certain types of polythene bags are not suitable for hybridisation with biotin-labelled probes as they lead to high backgrounds. Layflat polythene tubing (Trans Atlantic Supplies) and Sears' Boilable Cooking Pouches are both suitable.

(iv) Since very low background signals are obtained with biotinylated probes, the concentration of probe can be increased to 250—750 ng/ml in the hybridisation solution. This has the additional advantage of allowing short hybridisation times of 1—2 h.

10. DETECTION AND QUANTITATION OF HYBRIDS

The detection of hybrids involving probes labelled with non-radioactive markers is described in Chapter 2, Section 4.3. Here we shall consider only radioactive hybrids.

10.1 Detection

For detecting [³²P]hybrids, autoradiography is the most commonly used technique. It is sensitive, gives good resolution and does not involve destruction of the filter.

If the filter is not to be re-screened or re-used, dry it at room temperature or under an infra-red lamp. Then expose it to X-ray film (e.g., Kodak X-Omat RP) at room

temperature in a light-proof cassette. The time of exposure will vary from several hours to 14 days depending on the level of radioactivity in the hybrids. As a rough guide, a dot containing 100 c.p.m. ^{32}P will give a good signal on X-ray film with an overnight exposure. If the radioactivity levels are low, use of an intensifying screen (e.g., Ilford fast tungstate or Fuji Mach II) increases the sensitivity of the film by 4- to 5-fold. The film is sandwiched between the filter and the intensifying screen in the cassette. Exposure is at −70°C because fluorescence reflected off the intensifying screen is prolonged at low temperatures. If two intensifying screens are used, the sensitivity of the film is enhanced 8−10 times. In this case the filter and film are sandwiched between the two intensifying screens. For low levels of radioactivity, the film can be pre-flashed and placed flashed side against the intensifying screen with the filter on top (Chapter 5, Table 3, step 15). As few as 5−10 c.p.m. per dot above background can be evaluated dependably by this adaptation of autoradiography.

If the filter is to be re-used (Section 13.1), it must not be allowed to dry otherwise the probe will bind irreversibly. The wet filter is covered in Clingfilm or Saranwrap or is inserted into a thin polythene sleeve before exposure to X-ray film. The filter should not be too wet or ice crystals will form when the cassette is placed at −70°C. This will distort the filter and could cause it to crack.

10.2 Quantitation of Hybridisation Signals

For many purposes it is sufficient to compare visually the intensity of hybridisation signal on an autoradiogram with that generated by a standard series of dots. The accuracy is better than 2-fold over a 100-fold range, taking into account both the intensity and diameter of the autoradiographic spots. For example, Figure 4 shows dot blots of genomic DNA from patients suffering from chronic myeloid leukaemia (CML) and from two CML cell lines, K562 and NALM-1, probed with the c-sis and c-abl oncogenes and a human immunoglobulin λ light chain variable gene sequence (IgV$_\lambda$). Visual comparison of the autoradiographic signals indicates that cell line K562 contains about four times more copies of the c-abl and IgV$_\lambda$ genes than the other cell lines.

For more accurate quantitation, densitometry can be used. This is a very simple and sensitive procedure; as little as 5−10 c.p.m. above background can be evaluated reliably. It is the best method of quantitation when the amount of radioactivity in hybrids is low. A scan is made of a series of standard dots and of the unknown samples. The area under the peaks is integrated, either electronically or the peaks can be cut from paper traces and weighed. The weight of the paper is a measure of the autoradiographic signal. A graph is then plotted of the weight of (or area under) the standard peaks against the known amount of nucleic acid on the filter. The concentration of the probe must be in excess over that on the filter and the autoradiograph should not be overexposed. An example of densitometric quantitation of a blot is given in Figure 5. Note that the curve relating intensity of signal (area under the peak) to the amount of RNA in the dot is only linear for a restricted range of amounts of filter-bound RNA. So, for the probe used in Figure 5, quantitation can be carried out only over the range 0−6 μg RNA per dot since beyond this the filter-bound sequences are in excess of this probe. The curve in Figure 5c does not reach a plateau in the range analysed because the size of the dots is not uniform.
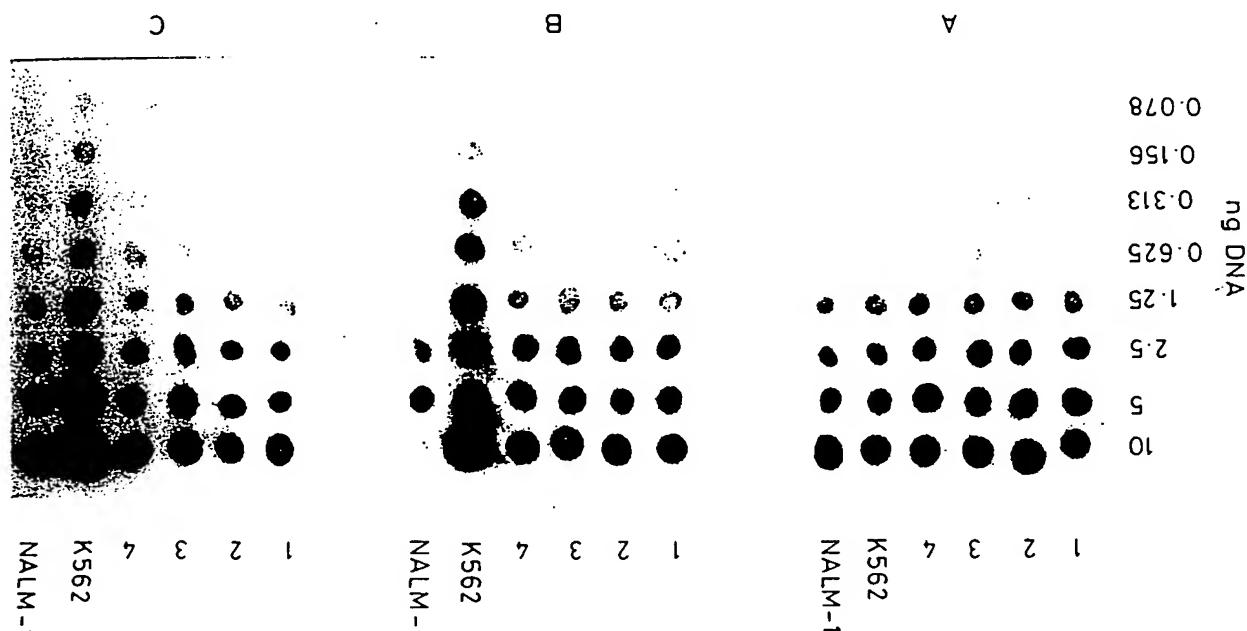
Figure 4. Genomic DNA dot blots. Replicate nitrocellulose filters containing the indicated amounts of genomic DNA from a dilution series were probed with (A) ^{32}P-labelled c-sis oncogene, (B) c-abl oncogene and (C) Ig V$_\lambda$ DNA. Lanes 1−4 contained DNA from peripheral blood of chronic myeloid leukaemia (CML) patients. 5 and 6 contained DNA from CML cell lines K562 and NALM-1, respectively.

Figure 6. Plasmid DNA dot blots. Duplicate filters contained cloned recombinant plasmid cDNAs derived from mRNA of a patient suffering from acute non-lymphocytic leukaemia (ANLL). Probes were [32P]cDNA derived from unfractionated mRNA of (a) ANLL and (b) chronic lymphocytic leukaemia (CLL) patients. The DNA bound to the filter was in significant excess over the probe. Data from M. Warnock, with permission.

If the hybrids are sufficiently radioactive, the dots can be cut out and counted in a liquid scintillation counter. This means, of course, that the filter cannot be re-hybridised.

10.3 Quantitative Analysis of Nucleic Acid Complexity

As we have seen, the rate of hybridisation is inversely proportional to the complexity of nucleic acid for both solution hybridisation and nucleation-limited filter hybridisation although not for diffusion-limited filter hybridisation (Section 2). Reassociation kinetics in solution have been used extensively to analyse the complexity of DNA and RNA populations and it might be supposed that nucleation-limited filter hybridisation could be used for a similar purpose. However, filter hybridisation is not suitable for quantitative studies of complexity. This is because the rate of filter hybridisation is so low that it is difficult to obtain C_0t values high enough for single copy sequences to hybridise (see Section 2).

10.4 Measurement of Relative Abundance of RNA Transcripts

For high and medium-abundance classes, dot blot hybridisation can be used to measure the relative prevalences of different mRNA species (44,45). Cloned recombinant cDNAs are applied in dots to filters and hybridised with either labelled mRNA or the cDNA derived from it. Filter-bound DNA is in excess so the extent of hybridisation is a measure of the concentration of the cloned cDNA sequence in the mRNA probe. [The extent of reaction is a reproducible characteristic of each clone and not a function of the cloned insert length, at least between the limits of 400–1500 nucleotides tested (44,45)]. It is estimated that a clone must be represented to a level of at least 0.1% of the mass of mRNA to be detected (44,46). This is probably true for optimal reaction conditions, but in practice the lower limit is more likely to be nearer 0.5%. An example is shown in *Figure 6*. Recombinant cDNA clones were constructed using mRNA from a patient

Figure 5. Quantitation of RNA dot blots. (a) Serial dilutions of RNA from patient 2 in *Figure 3* were prepared and applied in duplicate (A,B) to a nitrocellulose filter. The filter was hybridised with the Ha-ras-1 probe (*Figure 3*). (b) Densitometric scan made across lane A. (c) The relationship between the amount of RNA in each dot and the area under the densitometric peak for each dot.

As expected, only one of the separated strands of DNA hybridises with the mRNA (*Figure 7a*). The number of transcripts per ng mRNA is then determined from the slope of the line using the relationship:

$$T = \frac{fN}{LS \times 350 \times 10^9}$$

where T is the number of transcripts per ng mRNA, f is the slope of the titration curve (c.p.m./ng mRNA). N is Avogadro's Number (number of molecules/ mol). L is the length of hybrid (i.e., cDNA insert length in nucleotides). S is the specific activity of the labelled cDNA (c.p.m./ng). and 350 is the average molecular weight of a ribonucleotide.

The appropriate values for clones A2. C1 and D2 of *Figure 7a* are:

Clone	Specific activity (S)	Length of insert (L)	Slope (f)	No. of transcripts/ng mRNA (T)
A2	2.0×10^4	800	4.7	5×10^5
C1	1.5×10^4	1000	26.3	3×10^6
D2	1.0×10^4	750	89.4	2×10^7

Using these values of T, a graph is plotted of the number of transcripts against the radioactivity in the hybrids on the dot blots. A linear relationship is obtained (*Figure 7b*). From this graph, the number of transcripts of any other recombinant on the dot blot matrix can easily be obtained.

11. DETERMINATION OF OPTIMAL REACTION CONDITIONS

11.1 Buffer Composition and Temperature

To determine the optimal reaction conditions, prepare replicate dot blots. Hybridise some under different hybridisation conditions, keeping the washing conditions constant. and monitor the effects. Then hybridise other dot blots under optimal hybridisation conditions and vary the washing conditions. For an extensive analysis of the effects of altering conditions on the hybrids formed, the reader is referred to references 3–5 and 43. The following is a rough guide.

(i) Reaction conditions which favour the detection of well-matched hybrids involve high temperatures of hybridisation (65–68°C in aqueous solution and 42°C in 50% formamide) combined with washing at high temperatures (5–25°C below T_m) and at low salt concentrations (0.1 x SSC).

(ii) To detect poorly-matched hybrids, filters should be hybridised in solutions containing formamide (20–50%) at 35–42°C but washed at high salt concentrations at an intermediate temperature (e.g., 2–6 x SSC at 40–60°C). Again. conditions may have to be determined empirically. It should be remembered that both closely-related and distantly-related sequences will be detected under these conditions.

Figure 7. Prevalence analysis of mRNA transcripts. (a) Titration of cDNA clones with poly (A)+ mRNA. Complementary strands (open and closed circles) of [¹²⁵I] cDNA recombinant plasmids A2. C1. and D. (*Figure 6*) are hybridised separately in solution with the indicated amounts of poly(A)+ RNA from ANLI cells. Reactions are carried out to 20 x $C_0t_{1/2}$ calculated with respect to the labelled DNA which is in sequence excess. The RNA-DNA hybrids formed in the reaction are analysed by resistance to nuclease S1. The slope of the lines was used to determine the number of transcripts per ng RNA. (Section 10.2). (b) Relationship between [³²P]cDNA dot blot hybridisation (*Figure 6*) and the number of transcripts per ng RNA for clones A2. C1 and D2.

suffering from acute non-lymphocytic leukaemia (ANLL). These were screened with [³²P]cDNA synthesised using unfractionated mRNA from an ANLL patient as template and with [³²P]cDNA complementary to unfractionated mRNA of a patient suffering from chronic lymphocytic leukaemia (CLL). Recombinant clones representing mRNAs common to the two diseases (e.g., B5, C1), and their relative abundance, can be easily identified and distinguished from those apparently specific to ANLL (e.g., D2, D3, D4, D5).

By using a calibration curve, the extent of hybridisation of [³²P]cDNA to each filter-bound recombinant can be used to determine the actual number of transcripts of mRNA (47,48). From the dot blot results (*Figure 6*), at least three cDNA clones whose representation in mRNA differs in abundance over a wide range are selected (e.g., clones A2. C1 and D2). Separated single strands of these recombinant DNAs are radiolabelled and hybridised separately in solution to different amounts of mRNA. The concentration of reactants is adjusted such that recombinant DNAs are in sequence excess. The reactions are carried out to about 20 x $C_0t_{1/2}$. The radioactivity in nuclease S1-resistant hybrids is determined and plotted against the amount of mRNA added to the reaction.

M.L.M. Anderson et al., B.D. Young

(iii) To distinguish between closely- and distantly-related members of the same family, conditions must be found which are permissive for some sequences and stringent for others. As we have already seen (Section 5), distant homologies are best detected when the ratio of rate constants for hybridisation of cross-hybridising to self-hybridising species is high, whereas closely-related species are most easily detected when the ratio is low. In practical terms, this means that for distantly-related hybrids low temperatures are used, whereas for closely-related hybrids high temperatures are best. The time of incubation is very important since the effective discrimination between closely- and distantly-related hybrids is highest with very short times of incubation and deteriorates very rapidly thereafter (Section 5). If short incubations do not give a sufficiently high hybridisation signal, then longer times can be used with excess filter-bound nucleic acid.

11.2 Time Period of Incubation

It is difficult to give a precise time period for hybridising filters. Filter hybridisations tend not to go to completion. As described above, the rate of hybridisation on filters is about 10 times slower than that for solution hybridisation of the same DNAs (9), so it is difficult experimentally to reach the very high C_0t values required for complete hybridisation. Prolonged incubation does not necessarily increase the extent of hybridisation because:

(i)   more and more probe reassociates
(ii) at the high temperatures involved, sequences leach off the filter if they are not covalently bound
(iii) the probe is gradually degraded.

Addition of formamide to the hybridisation solution allows lower temperatures to be used and thus incubation times can be extended, but there is still the problem of probe reassociation unless a single-stranded probe is used. Furthermore, steric constraints prevent all the bound sequences being nucleated effectively. Under optimal conditions, no more than 80% of a single sequence probe appears in hybrids (9) and so an even smaller proportion of more complex probes will be hybridised.

In practice, for double-stranded DNA probes, there is no need to proceed for longer than to allow the probe in solution to achieve $1-3 \times C_0t_{1/2}$. After incubation for 3 x $C_0t_{1/2}$, the amount of probe available for additional hybridisation to sequences on the filter is negligible. The following useful guideline is taken from ref. 46. In 10 ml hybridisation solution, 1 μg denatured double-stranded probe with a complexity of 5 kb will reach $C_0t_{1/2}$ in 2 h. To determine the number of hours (n) needed to achieve $C_0t_{1/2}$ for renaturation of any other probe, the appropriate values can be substituted in the following equation:

$$n = \frac{1}{X} \times \frac{Y}{5} \times \frac{Z}{10} \times 2$$

where X is the weight of the probe added (in μg), Y is its complexity (which for most probes is proportional to the length of the probe in kb) and Z is the volume of the reaction (in ml).
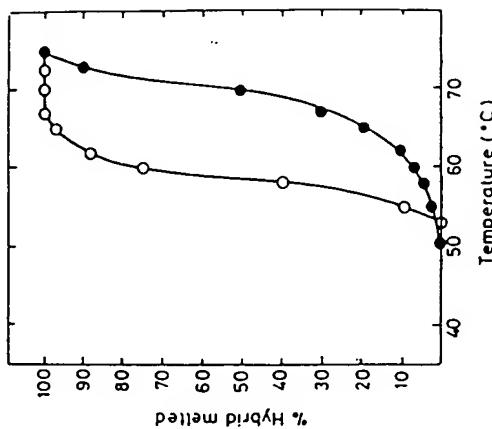

Figure 8. Melting profile of dot hybrids. Replicate filters containing dots of linearised DNA are hybridised to homologous 32P-labelled DNA (○) or RNA (●). Hybridisation, in a formamide-containing solution, is carried out to 20 x $C_0t_{1/2}$. The DNA bound to the filter is present in significant excess. After hybridisation each filter is washed in the same formamide-containing buffer as used for hybridisation, with stepwise increases in temperature. The filters are counted by Cerenkov counting between each washing step to determine the percentage of hybridised [32P]DNA or [32P]RNA which has eluted (melted) at each temperature.

For quantitative dot blots (e.g., Figure 6) the time period of incubation, the concentration of probe and the amount of nucleic acid bound to the filter should all be adjusted such that there is low fractional hybridisation of both the probe and filter-bound sequences. This is to ensure that the proportion of the probe hybridising increases linearly with time.

Factors affecting the selection of a time of incubation suitable for discrimination between related sequences are discussed in Section 5.2 [see also Section 11.1 (iii)].

12. MEASUREMENT OF T_m

The DNA or RNA sample is applied in a dot to a filter and hybridised with a labelled DNA or RNA probe as appropriate. It is then washed at the same temperature in a solution of the same composition as was used for the hybridisation. This removes unhybridised and non-specifically-bound probe. For melting of the hybrid, the filter is incubated for 10-15 min in a small volume of the same buffer at progressively higher temperatures. The buffers used must be pre-heated to the required temperature. Between melting steps, the filter is counted by Cerenkov counting, making sure that it does not dry out at any stage in order to prevent irreversible binding of the probe. Replicate filters can be hybridised at different criteria or washed in solutions of different ionic strength. Figure 8 shows the melting temperature profiles to be expected from homologous DNA-DNA and RNA-DNA hybridisations. Note that in formamide-containing

solutions RNA-DNA hybrids are more stable than the corresponding DNA-DNA hybrids and so dissociate (melt) at higher temperatures. The exact temperature difference will depend on the base composition of the nucleic acids involved, the size of the probe and the degree of relatedness of the two hybridising species. In the example given in *Figure 8*, the difference in T_m is 11°C.

For a single nucleic acid species hybridising to itself, the hybrid melts over a very narrow temperature range and the T_m is the same irrespective of the incubation temperature. T_i. However, when the same nucleic acid is probed with a complex mixture of sequences which have varying degrees of relatedness, the T_m profile depends on the reaction conditions. For hybrids formed at low criterion ($T_m - 25$°C), the melting profile is broad because both well-matched and poorly-matched hybrids are formed. They melt at different temperatures, so the overall melting profile, which is a composite of the contributions of all the hybrids, will reflect this. At high criterion ($T_m -8$°C), only hybrids with a high degree of homology form so they melt over a very narrow temperature range. The melting profile is also broad when variable length probes are used. This is most apparent at short average lengths of hybridised probe in accordance with the empirical relationship:

$$T_n - T_m = \frac{650}{L}$$

where L is the length of the probe in nucleotides, T_m is the melting temperature of the short hybrid and T_n is the melting temperature of long DNA molecules (17).

The procedure of stepwise melting of hybrids described above for T_m measurement can be extended to investigate the degree of relatedness of different sequences and can be applied to many samples at once. An array of dots on a single filter are hybridised to a labelled probe and the extent of similarity to the probe is evaluated by stepwise melting and autoradiography (3). The more mismatched a hybrid is, the lower the temperature at which it will melt. Thus differences in the intensity of signal of the dots can be interpreted in terms of the degree of relatedness of the different sequences.

13. RE-USE OF FILTERS AND PROBES

13.1 Filters

In many cases, re-probing the same filters with a series of different probes yields valuable information. Filters can be re-probed several times — the exact number being dependent on the type of filter and the incubation conditions to which the filter has already been exposed. Nitrocellulose filters which have been exposed to high temperatures for hybridisation and washing can be used about two to four times before falling apart. Filters exposed to the less harsh conditions of hybridisation at lower temprature in the presence of formamide can be used many more times. Nylon filters can be used indefinitely without disintegration and so, because of their superior durability, they are preferred to nitrocellulose for multiple probings.

Before re-probing a filter with a new probe, it is first necessary to strip off the old probe and to monitor that the treatment has been effective. This can be achieved for both DNA and RNA dot blots as follows:

(i) Transfer the damp filter to a plastic box containing 200 ml of 5 mM Tris-HCl, pH 8.0, 0.1 mM EDTA, 0.05% sodium pyrophosphate, 0.1% Denhardt's solution (see *Table 3*, footnote c for composition) at 65°C. Incubate for 1 h with gentle agitation.

(ii) Discard the wash buffer and repeat step (i).

(iii) Check by autoradiography as described in Section 10.1 that the probe has been removed.

An alternative procedure for DNA dot blots only is given below.

(i) Wash the filter twice, for 10 min each wash, in 50 mM NaOH at room temperature.

(ii) Wash and neutralise the filter by incubating it (5 min each time) in five changes of TE buffer, pH 7.5 (see *Table 1*, footnote a) at room temperature.

This procedure cannot be carried out successfully for RNA dot blots since the NaOH hydrolyses the filter-bound RNA.

Unfortunately there are several potential problems with re-using filters:

(i) Loss of sensitivity. Prolonged use of both nylon and nitrocellulose filters leads to gradual reduction in sensitivity through loss of filter-bound nucleic acid (Section 6.1). This is not a problem if the nucleic acid has been covalently bound to the filter matrix.

(ii) Irreversible binding of the previous probe if the filter was allowed to dry.

(iii) Incomplete removal of the previous probe even if the filter was kept wet throughout its use. This can give misleading results if single-stranded tails of probe remaining on the filter are complementary to sequences of the new probe. For example, if the first probe is a recombinant DNA (vector and insert) and the second is also a recombinant DNA with a different insert, hybridisation may occur through plasmid sequences and this will obscure the hybridisation of the second probe insert sequences. Therefore, it is recommended that all inserts are excised from vectors before labelling as probes.

13.2 Probes

Normally only a small fraction of the probe is used up during hybridisation, so probes can be re-used until they are degraded or have decayed to too low a specific activity. To re-use the probe (now in the hybridisation solution) it must be denatured again by heating to a temperature above its T_m. For aqueous solutions, this can be done by incubating in a boiling water bath for 10 min. For formamide-containing solutions, heat at 70°C for 30 min. The newly-denatured probe can now be added to a second filter which has been pre-hybridised under standard conditions.

14. PROBLEMS

Most hybridisation experiments using filter-bound nucleic acids employ radioactive probes and so problems of only this type of investigation are covered here.

(i) The autoradiograph of the filter is black all over:
   (a) At some stage during hybridisation or washing, the filter was allowed to

dry. It will probably be necessary to strip the filter (Section 13.1) and re-hybridise.

  (b) The probe is 'dirty'. It may be contaminated by traces of agarose. Either re-purify the nucleic acid from which the labelled probe was derived and prepare a new probe, or pass the labelled probe through a nitrocellulose filter which has been pre-treated in 10 x Denhardt's solution or through a mini NACS column (BRL).

  (c) An inappropriately low hybridisation and/or washing temperature was used.

(ii) The autoradiograph of the filter is black in parts:

  (a) Part of the filter dried out: see above.

  (b) The filter was handled with bare hands. Grease marks from fingers trap probe. Wear disposable plastic gloves in future.

(iii) The autoradiograph has black dots in random locations:

  (a) The unincorporated precursors were not completely removed from the probe. See correct procedure in relevant section of Chapter 2.

  (b) Air bubbles were not completely removed from the bag during hybridisation. (This may not matter if a shaking water bath is used, but the effect may be quite troublesome if the bag is not agitated.)

  (c) Dust or dirt on the filter. Filter all solutions before use in future.

(iv) The signal is lower than expected:

  (a) Was the correct binding procedure used? Nitrocellulose and nylon filters use different binding protocols (see Section 6.1).

  (b) The probe was degraded. This is most likely to happen with RNA probes.

  (c) The double-stranded probe was not denatured (see Section 8.2.2).

  (d) The hybridisation and/or washing conditions were too stringent so that the hybrids either did not form at all, or were dissociated.

  (e) The specific activity of the probe was too low.

  (f) The hybridisation time was too short.

  (g) The filter was not exposed to film for long enough.

(v) A 'negative' effect is obtained, that is, the background of the autoradiograph is black with clear dots. Too high a concentration of [^{32}P]probe was used.

(v) The filter fell apart. This is most likely to occur with nitrocellulose filters.

  (a) During binding of DNA to the filter, the alkali was not properly neutralised thus making the filter yellowish in colour and very brittle.

  (b) After repeated use, the filter becomes brittle despite correct procedures. Prepare new filters.

15. ACKNOWLEDGEMENTS

16. REFERENCES

1. Gillespie,D. and Spiegelman,S. (1965) *J. Mol. Biol.*, **12**, 829.
2. Britten,R.J. and Kohne,D.E. (1968) *Science (Wash.)*, **161**, 529.
3. Sun,G.K., Kafatos,F.C., Jones,C.W., Koehler,M.D., Efstratiadis,A. and Maniatis,T. (1979) *Cell*. ...
4. Beltz,G.A., Jacobs,K.A., Eickbush,T.H., Cherbas,P.T. and Kafatos,F.C. (1983) *Methods in Enzymology*, vol. 100, Wu,R., Grossman,L. and Moldave,U. (eds.), Academic Press, NY, p. 266.
5. Kafatos,F.C., Jones,C.W. and Efstratiadis,A. (1979) *Nucleic Acids Res.*, **7**, 1541.
6. Wetmur,J.G. (1971) *Biopolymers*, **10**, 601.
7. Lee,C.H. and Wetmur,J.G. (1972) *Biopolymers*, **11**, 549.
8. McCarthy,B.J. and McConaughy,B.L. (1968) *Biochem. Genet.*, **2**, 37.
9. Flavell,R.A., Birfelder,E.J., Sanders,J.P. and Borst,P. (1974) *Eur. J. Biochem.*, **47**, 535.
10. Birnsteil,M.L., Sells,B.H. and Purdom,I.F. (1972) *J. Mol. Biol.*, **63**, 21.
11. Flavell,R.A., Borst,P. and Birfelder,E.J. (1974) *Eur. J. Biochem.*, **47**, 545.
12. Wetmur,J.G. and Davidson,N. (1968) *J. Mol. Biol.*, **31**, 349.
13. Marmur,J.G. and Doty,P. (1961) *J. Mol. Biol.*, **3**, 584.
14. Bonner,T.I., Brenner,D.J., Neufield,B.R. and Britten,R.J. (1973) *J. Mol. Biol.*, **81**, 123.
15. Howley,P.M., Israel,M.F., Law,M-F. and Martin,M.A. (1979) *J. Biol. Chem.*, **254**, 4876.
16. Casey,J. and Davidson,N. (1977) *Nucleic Acids Res.*, **4**, 1539.
17. Britten,R.J., Graham,D.E. and Neufield,B.R. (1974) in *Methods in Enzymology*, Vol. 29E, Grossman,L. and Moldave,K. (eds.), Academic Press, NY, p. 363.
18. Wetmur,J.G. (1975) *Biopolymers*, **14**, 2517.
19. Wahl,G.M., Stern,M. and Stark,G.R. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 3683.
20. Schildkraut,C. and Lifson,S. (1965) *Biopolymers*, **3**, 195.
21. McConaughy,B.L., Laird,C.D. and McCarthy,B.J. (1969) *Biochemistry (Wash.)*, **8**, 3289.
22. Marmur,J. and Doty,P. (1962) *J. Mol. Biol.*, **5**, 109.
23. Dove,W.F. and Davidson,N. (1962) *J. Mol. Biol.*, **5**, 467.
24. Schmeckpeper,B.J. and Smith,K.D. (1972) *Biochemistry (Wash.)*, **11**, 1319.
25. Hyman,R.W., Brunovskis,I. and Summers,W.C. (1973) *J. Mol. Biol.*, **77**, 189.
26. Yung,R.C., Young,A. and Wu,R. (1980) *J. Virol.*, **34**, 416.
27. Haas,M., Vogt,M. and Dulbecco,R. (1972) *Proc. Natl. Acad. Sci. USA*, **69**, 2169.
28. Church,G.M. and Gilbert,W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1991.
29. Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503.
30. Nagamine,Y., Sentenac,A. and Fromageot,P. (1980) *Nucleic Acids Res.*, **8**, 2453.
31. Parnes,J.R., Velan,B., Felsenfeld,A., Ramanathan,L., Ferrini,U., Apella,E. and Seidman,J.G. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 2253.
32. Thomas,P. (1983) in *Methods in Enzymology*, Vol. 100, Wu,R., Grossman,L. and Moldave,U. (eds.), Academic Press, NY, p. 255.
33. McMaster,G.K. and Carmichael,G.G. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 4835.
34. Bailey,J.M. and Davidson,N. (1976) *Anal. Biochem.*, **70**, 75.
35. Lehrach,H., Diamond,J., Wozney,J.M. and Boedker,H. (1977) *Biochemistry (Wash.)*, **16**, 4743.
36. Bantle,J.A., Maxwell,I.H. and Hahn,W.E. (1976) *Anal. Biochem.*, **72**, 413.
37. Thomas,P. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5201.
38. Renz,M. and Kurz,C. (1984) *Nucleic Acids Res.*, **12**, 3435.
39. Langer,P.R., Waldrop,A.A. and Ward,D.C. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 6633.
40. Leary,J.J., Brigati,D.J. and Ward,D.C. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 4045.
41. Denhardt,D.T. (1966) *Biochem. Biophys. Res. Commun.*, **23**, 641.
42. Singh,L. and Jones,K.W. (1984) *Nucleic Acids Res.*, **12**, 5627.
43. Meinkoth,J. and Wahl,G. (1984) *Anal. Biochem.*, **138**, 267.
44. Williams,J.G. and Lloyd,M.M. (1979) *J. Mol. Biol.*, **129**, 19.
45. Dworkin,M. and Dawid,I.B. (1980) *Dev. Biol.*, **76**, 435.
46. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning, A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY.
47. Lasky,L.A., Lev,Z., Xin,J-H., Britten,R.J. and Davidson,E.H. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5317.
48. Xin,J-H., Brandhorst,B.P., Britten,R.J. and Davidson,E.H. (1982) *Dev. Biol.*, **89**, 527.

Comparison of Biosequences

TEMPLE F. SMITH

Northern Michigan University, Marquette, Michigan 48955

AND

MICHAEL S. WATERMAN

Los Alamos Scientific Laboratory, Los Alamos, New Mexico 87545

Homology and distance measures have been routinely used to compare two biological sequences, such as proteins or nucleic acids. The homology measure of Needleman and Wunsch is shown, under general conditions, to be equivalent to the distance measure of Sellers. A new algorithm is given to find similar pairs of segments, one segment from each sequence. The new algorithm, based on homology measures, is compared to an earlier one due to Sellers.

1. DISTANCE AND SIMILARITY

Both distance and similarity measures have been designed for the comparison of pairs of biological molecules. The basis of such comparisons is the information from the biochemist as to the linear sequence of elements comprising such macromolecules, as the DNA of the gene. These distance and/or similarity measures have been used by the biologist to obtain information about processes of molecular evolution. The simplest and most fundamental of these are the point mutation (the conversion of one sequence element into another) and the insertion or deletion of sequence elements.

In 1970 Needleman and Wunsch [1] introduced their homology (similarity) algorithm. From a mathematical viewpoint, their work lacks rigor and clarity. But their algorithm has become widely used by the biological community for sequence comparisons.

The two molecules under consideration will be denoted by $a = a_1 a_2 \ldots a_n$ and $b = b_1 b_2 \ldots b_m$. The basic problem is to find the alignment of a and b with the highest similarity. To be specific, we define an alignment of a and b by $A(a, b) = \{(a_{i_1}, b_{j_1}), (a_{i_2}, b_{j_2}), \ldots : 1 \leq i_1 < i_2 < \cdots \leq n, 1 \leq j_1 < j_2 < \cdots \leq$

$m\}$. Each $a_k (b_k)$ not appearing in the subsequence $a_{i_1} a_{i_2} \ldots (b_{j_1} b_{j_2} \ldots)$ will be considered an insertion or deletion, depending on the point of view. For display of an alignment, the null element Δ will be inserted in the sequences to indicate insertions/deletions. Thus, the alignment $A = \{(a_1, b_3), (a_2, b_4), (a_4, b_5)\}$ for $a = a_1 a_2 a_3 a_4$ and $b = b_1 b_2 b_3 b_4 b_5$ is displayed

$$
\begin{array}{ccccc}
\Delta & \Delta & a_1 & a_2 & a_3 & a_4 \\
b_1 & b_2 & b_3 & b_4 & \Delta & b_5
\end{array}
$$

Frequently, insertions/deletions of length greater than one are used. Then the display becomes

$$
\begin{pmatrix} \Delta \\ b_1 \end{pmatrix} \begin{array}{|cccc} a_1 & a_2 & a_3 & a_4 \\ b_3 & b_4 & \Delta & b_5 \end{array}
$$

Similarity measures are based on two weight functions. The first, $s(a_i, b_j)$, measures the degree of "similarity" between two elements a_i, b_j. For ease of notation, let (a_i, b_j) be known as a match of type k if it is assigned weight α_k:

$$
s(a_i, b_j) = \alpha_k.
$$

The other function necessary is $w_k \geq 0$, the weight assigned to an insertion/deletion of length k. Now let λ_k be the number of matches of type k and Δ_k be the number of insertions/deletions of length k. The similarity measure between a and b is then

$$
S(a, b) = \max_A \left(\sum \alpha_i \lambda_i - \sum w_k \Delta_k \right).
$$

The first theorem provides an algorithm to calculate S. The statement in Needleman and Wunsch [1] is less general, unclearly stated, and does not have a proof.

THEOREM 1. *Let* $S_{0j} = -w_j$, $S_{i0} = -w_i$, $0 \leq j \leq m$, $0 \leq i \leq n$. *If* $S_{ij} = S(a_1 a_2 \ldots a_i, b_1 b_2 \ldots b_j)$, *then*

$$
S_{ij} = \max \Big\{ S_{i-1, j-1} + s(a_i, b_j), \max_{k \geq 1} (S_{i, j-k} - w_k),
$$
$$
\max_{l \geq 1} (S_{i-l, j} - w_l) \Big\}.
$$

Proof. Let A be an optimal alignment for $a_1 \ldots a_i, b_1 \ldots b_j$. There are three cases: (i) a_i is matched with b_j. Then the remaining sequences

484.   SMITH AND WATERMAN

$a_1 \ldots a_{i-1}, b_1 \ldots b_{j-1}$ must be optimally aligned and S_{ij} equals

$$S_{i-1,j-1} + s(a_i, b_j).$$

(ii) a_i is a member of an insertion/deletion of length k and S_{ij} equals

$$S_{i-k,j} - w_k.$$

(iii) b_j is a member of an insertion/deletion of length l and S_{ij} equals

$$S_{i,j-l} - w_l.$$

T. F. Smith and S. M. Ulam realized that for the purposes of taxonomic tree construction, it would be appropriate to have a metric D defined for biological sequences. The mathematical community was made aware of this problem by Ulam. P. H. Sellers learned of this problem and solved it in 1972 [2].

As in our discussion of similarity we let (a_i, b_j) be known as a match of type k if it is assigned weight

$$d(a_i, b_j) = \beta_k.$$

Here, d is a "distance" between a_i and b_j, and d is required to be a metric on the set of sequence elements. Sellers only allowed insertions/deletions of length one, but the generalization to insertions/deletions of length k was later made by Waterman et al. [6]. Deletions of length k are assigned weight $x_k \geq 0$. The distance measure between a and b is then

$$D(\mathbf{a}, \mathbf{b}) = \min_A \left\{ \sum_i \beta_i \lambda_i + \sum_k x_k \Delta_k \right\}.$$

The next theorem was given in [6] and generalizes the work of Sellers [2]. The proof follows the general lines of Theorem 1, but the inclusion of longer insertions/deletions is more difficult.

THEOREM 2. *If* $x_1 \leq x_2 \leq \cdots$ *and* d *is a metric on the set of sequence elements, then* D *is a metric on the set of sequences. Let* $D_{i0} = x_i$, *and* $D_{0j} = x_j$ *for* $0 \leq i \leq n, 0 \leq j \leq m$. *If* $D_{ij} = D(a_1 a_2 \ldots a_i, b_1 b_2 \ldots b_j)$, *then*

$$D_{ij} = \min \left\{ D_{i-1,j-1} + d(a_i, b_j), \min_k (D_{i,j-k} + x_k), \right.$$
$$\left. \min_l (D_{i-l,j} + x_l) \right\}.$$

Until recently [5], it was not known whether the Needleman–Wunsch algorithm and the Sellers algorithm were equivalent or not. This was largely

COMPARISON OF BIOSEQUENCES   485

due to the differences in the way the algorithms were formulated and to the question not being clearly stated. Here the two algorithms are defined to be equivalent if given the weights for one algorithm there is a choice of weights for the second algorithm such that the set of alignments achieving the maximum value for Needleman–Wunsch is equal to the set of alignments achieving the minimum value for Sellers. The following theorem is contained in [5].

THEOREM 3. *The Needleman–Wunsch similarity algorithm is equivalent to the Sellers algorithm. The equivalence is established by setting*

$$\beta_i = \max_j (\alpha_j) - \alpha_i$$

and

$$x_k = k/2 \max_j (\alpha_j) + w_k.$$

Proof. As above, let λ_i = the number of matches of type i and Δ_k = the number of deletions of length k. The proof is based on the observation that

$$n + m = 2 \sum_i \lambda_i + \sum_k k \Delta_k.$$

To be specific, suppose a Needleman–Wunsch algorithm is given. Let

$$\alpha_M = \max_i \alpha_i$$

and

$$\beta_i = \alpha_M - \alpha_i.$$

Then,

$$S = \max_A \left\{ \sum_i \alpha_i \lambda_i - \sum_k w_k \Delta_k \right\}$$

$$= \max_A \left\{ \alpha_M \sum_i \lambda_i - \sum_i \beta_i \lambda_i - \sum_k w_k \Delta_k \right\}.$$

But

$$\sum_i \lambda_i = \frac{n+m}{2} - \sum_k \frac{k}{2} \Delta_k.$$

SMITH AND WATERMAN

486

so

$$S = \max_A \left\{ a_M \frac{n+m}{2} - \sum_i \beta_i \lambda_i - \sum_k \left(\frac{a_M k}{2} + w_k \right) \Delta_k \right\}$$

$$= a_M \frac{n+m}{2} - \min_A \left\{ \sum_i \beta_i \lambda_i + \sum_k \left(\frac{a_M k}{2} + w_k \right) \Delta_k \right\}.$$

Therefore, the algorithms are equivalent if

$$\beta_i = a_M - a_i$$

and

$$x_k = \frac{a_M k}{2} + w_k.$$

Of course, not all choices of similarity a_i will induce a metric d on the sequence alphabet. But the cases of interest are all included. For example, the simplest cases have

$$s(a,b) = 0 \quad \text{if } a \neq b,$$
$$= 1 \quad \text{if } a = b$$

and

$$d(a,b) = 1 \quad \text{if } a \neq b,$$
$$= 0 \quad \text{if } a = b.$$

While $d = 1 - s$, we still must use $x_k = k/2 + w_k$.

2. MAXIMUM SIMILARITY SEGMENTS

Frequently two sequences which are, overall, no closer together than expected by random will have segments which are quite similar. For a sequence $a = a_1 a_2 \ldots a_n$, a segment is defined to be a subsequence $a_i a_{i+1} \ldots a_j$ where $1 \leq i \leq j \leq n$. After these similar segments are located, it is the task of a biologist to access their significance.

This problem of locating similar patterns within two sequences has been worked on by Sellers [3]. His approach is via distance measures and is quite involved. Our own approach makes use of similarity measures and is simpler. See [4] for an announcement of this result. The connection between these two approaches is discussed in the section following this one.

COMPARISON OF BIOSEQUENCES

487

As before, $a = a_1 a_2 \ldots a_n$ and $b = b_1 b_2 \ldots b_m$ are two biological sequences. We are given a similarity measure $s(a,b)$ between two sequence elements and a deletion weight w_k for deletions of length k. Define H_{ij} to be the maximum similarity of two segments that end in a_i and b_j, or zero, whichever is larger.

$$H_{ij} = \max\left(0, S(a_x a_{x+1} \ldots a_i, b_y b_{y+1} \ldots b_j): 1 \leq x \leq i \text{ and } 1 \leq y \leq j\right).$$

Zero arises from the view that negative values of H_{ij} represent less similar alignments than no association between the segments.

THEOREM 4. *Set $H_{i0} = H_{0j} = 0$ for $1 \leq i \leq n$ and $1 \leq j \leq m$. Then*

$$H_{ij} = \max\left\{ H_{i-1,j-1} + s(a_i, b_j), \max_{1 \leq k \leq i} (H_{i-k,j} - w_k),\right.$$
$$\left.\max_{1 \leq l \leq j} (H_{i,j-l} - w_l), 0 \right\}.$$

Proof. The proof is similar to that of Theorem 1. If the best segments have an alignment with a_i and b_j matched, the value of H_{ij} must be

$$H_{i-1,j-1} + s(a_i, b_j).$$

If a_i is a member of an insertion/deletion of length, k, H_{ij} must be equal to

$$H_{i-k,j} - w_k.$$

The case of b_j a member of an insertion/deletion is similar. Finally, H_{ij} equals zero if none of the above situations result in positive similarity.

The pair of segments with maximum similarity is found by first locating the maximum element of H. The other matrix elements leading to this maximum value are then sequentially determined with a traceback procedure ending with an element of H equal to zero. This procedure both identifies the segments and produces the corresponding alignment. The pair of segments with the next best similarity is found by applying the traceback procedure to the second largest element of H which is not associated with the first traceback and which has an alignment ending in a match.

Table 1 gives the matrix H for sequences AAUGCCAUUGACGG and CAGCCUCGCUUAG. Here $s(a,b) = 1$ if $a = b$, $s(a,b) = -\frac{1}{3}$ if $a \neq b$, and $w_k = 1 + k/3$. The maximum element is $H_{10,8} = 3.33$ and the corresponding members of the alignment are indicated by underlined traceback. The maximum similar segments are

        —GCCAUUG—
        —GCCAUCG—

which has five matches, one mismatch, and one deletion.

for all segment L and J satisfying $L \subset I \subset J \subset b$. One of his algorithms gives all segments $I \subset b$ most resembling a locally.

The final algorithm in [3] finds all segments $I \subset a$ and $J \subset b$ such that I most resembles J locally. Applying the equality $S + D = a_M n + m/2$, it can be seen that maximum similarity segments are found by this algorithm. However, the algorithm for maximum similarity segments runs in half the steps and is a great deal easier to implement. In addition, the Sellers algorithm does not contain an intrinsic optimality criterion for choosing segments I and J. Our algorithm orders the segments by the value of the similarity measure S.

REFERENCES

1. S. B. NEEDLEMAN AND C. D. WUNSCH, A general method applicable to the search for similarities in the amino acid sequence of two proteins, J. Mol. Biol. 48 (1970), 443–453.
2. P. H. SELLERS, On the theory and computation of evolutionary distances, SIAM J. Appl. Math. 26 (1974), 787–793.
3. P. H. SELLERS, The theory and computation of evolutionary distances: Pattern recognition, Algorithms, in press.
4. T. F. SMITH AND M. S. WATERMAN, The identification of common molecular subsequences, J. Mol. Biol. 147 (1981), 195–197.
5. T. F. SMITH, M. S. WATERMAN, AND W. M. FITCH, Comparative biosequence metrics. J. Mol. Evol., in press.
6. M. S. WATERMAN, T. F. SMITH, AND W. A. BEYER, Some biological sequence metrics, Adv. in Math. 20 (1976), 367–387.

TABLE I

Maximum Similarity Calculation

	C	A	G	C	U	C	G	C	U	U	C	A	G
A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00
A	0.00	1.00	0.66	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.66
U	0.00	0.00	0.66	0.00	1.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.66
G	0.00	0.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	0.66	0.00	0.66	1.00
C	0.66	0.00	0.00	1.33	0.33	0.00	0.66	0.66	0.33	0.33	2.00	0.33	0.33
C	2.00	0.00	0.00	3.00	1.00	0.33	1.00	1.66	1.66	0.33	1.33	0.00	0.00
C	0.66	2.00	0.00	1.00	1.33	1.33	0.66	0.66	2.66	1.00	0.66	0.33	0.00
A	2.00	1.00	0.00	0.33	1.33	1.00	0.66	0.66	2.66	1.33	0.66	1.33	0.00
U	0.66	0.33	0.66	0.33	2.33	1.00	2.00	0.66	0.66	2.33	0.66	1.00	1.00
U	0.33	0.33	1.66	1.00	2.66	0.66	2.00	1.66	1.66	2.66	1.33	1.33	2.66
G	0.00	1.00	0.00	2.00	2.33	0.66	3.00	1.66	1.66	1.33	1.66	2.33	2.00
A	1.00	0.00	0.66	0.66	1.00	1.00	3.00	1.00	1.00	1.33	2.66	2.33	2.00
C	0.00	1.00	1.00	2.00	0.66	1.00	1.66	2.66	2.66	2.66	1.66	1.00	2.00
G	0.66	0.00	1.00	0.66	0.66	0.33	0.33	2.66	2.66	2.33	2.66	2.33	2.00
G	0.00	0.00	1.66	0.33	0.33	1.33	0.66	1.33	1.33	1.33	1.33	2.33	2.00

3. CONCLUSION

The problem of locating segments of maximum similarity is that of finding segments I_0 and J_0 satisfying

$$\max_{I,J} S(I, J) = S(I_0, J_0)$$

when I is a segment of a and J is a segment of b. The proof of Theorem 3 shows

$$S(\mathbf{a}, \mathbf{b}) + D(\mathbf{a}, \mathbf{b}) = a_M(n + m)/2.$$

From this it might seem that the problem of segments of maximum similarity is equivalent to finding $\min_{I,J} D(I, J)$ but, if a and b have any elements in common, this minimum must be zero.

From this discussion, it is clear that the problem is more difficult to approach through the distance measures. This is the approach taken by Sellers [3]. He first defines a *most resembles* a segment I of b *locally* if

$$d(\mathbf{a}, I) \le d(\mathbf{a}, L)$$

and

$$d(\mathbf{a}, I) \le d(\mathbf{a}, J)$$